综合报告与论述

数 学*

数学是研究现实世界中数量关系和空间形式的，简单地说，数学是数和形的科学。

由于生活和劳动上的需求，即使是最原始的民族，也知道简单的计数，并由用手指或实物计数发展至用数字计数。在中国，至迟在商代，即已出现用十进制数字表示大数的方法；又至迟至秦汉之际，即已出现完满的十进位值制。在成书不迟于1世纪的《九章算术》中，已载有只有位值制才有可能的开平、立方的计算法则，并载有分数的各种运算以及解线性联立方程组的方法，还引入了负数概念。刘徽在他注解的《九章算术》（3世纪）中，还提出过用十进小数表示无理数平方根的奇零部分。但直至唐宋时期，欧洲则在16世纪S. 斯蒂文以后十进小数才获通用。在《九章注》中，刘徽又用圆内接正多边形的周长逼近圆周长，成为后世求圆周率更精确值的一般方法。虽然中国从来没有过无理数，或实数的一般概念，但在实质上，那时中国已完成了实数系统的一切运算法则与方法，这不仅在应用上不可缺，也为数学初期教育所不可少。至于继承了巴比伦、埃及、希腊文化的欧洲地区，则侧重于数的性质及这些性质间的逻辑关系的研究。早在欧几里得的《几何原本》中，即有素数的概念和素数个数无穷及整数唯一分解等论断。古希腊发现了有非分数的数，即现称的无理数。16世纪以来

---

* 本文摘自《中国大百科全书·数学卷》，中国大百科全书出版社，1988年。

由于解高次方程又出现了复数。在近代，数的概念更进一步抽象化并依据数的不同运算规律而对一般的数系统进行独立的理论探讨，形成数学中的若干不同分支。

开平方和开立方是解最简单的高次方程。在《九章算术》中，已出现解某种特殊形式的二次方程，发展至宋元时代，引进了“天元”（即未知数）的明确观念，出现了求高次方程数值解与求多至四个未知数的高次代数联立方程组的解的方法，通称为天元术与四元术。与之相伴出现的多项式的表达、运算法则以及消去方法，已接近于近世的代数学。在中国以外，9世纪阿拉伯的花拉子米的著作阐述了二次方程的解法，通常被视为代数学的鼻祖，其解法实质上与中国古代依赖于切割术的几何方法具有同一风格。中国古代数学致力于方程的具体求解，而导源于古希腊、埃及传统的欧洲数学则不同，一般致力于探究方程解的性质。16世纪时F. 韦达以文字代替方程系数，引入了代数的符号演算。对代数方程解性质的探讨，则从线性方程组导致行列式、矩阵、线性空间、线性变换等概念与理论的出现。从代数方程导致复数、对称函数等概念的引入以至伽罗瓦理论与群论的创立。近代极为活跃的代数几何，则无非是高次联立代数方程组解所构成的集体的理论研究。

形的研究属于几何学的范畴。古代民族都具有形的简单概念而往往以图画来表示，形之成为数学对象是由工具的制作与测量的要求所促成。规矩以作圆方，中国古代夏禹治水时即已有规、矩、准、绳等测量工具。《墨经》中对一系列的几何概念，有抽象概括，作出了科学的定义。《周髀算经》与刘徽《海岛算经》给出了用矩观天测地的一般方法与具体公式。在《九章算术》及刘徽注解的《九章算术》中，除勾股理论外，还提出了若干一般原理以解多种问题。例如出入相补原理以求任意多边形面积；阳马鳖臑的二比一原理（刘徽原理）以求多

面体体积；5 世纪祖暅提出“幂势即同则积不容异”的原理以求曲形体积特别是球的体积；还有以内接正多边形逼近圆周长的极限方法（割圆术）。但自五代（约 10 世纪）以后，中国在几何学方面的建树不多。中国几何学以测量与面积体积的量度为中心，古希腊的传统则重视形的性质与各种性质间的相互关系。欧几里得的《几何原本》，建立了用定义、公理、定理、证明构成的演绎体系，成为近代数学公理化的楷模，影响及于整个数学的发展。特别是平行公理的研究，导致了 19 世纪非欧几何的产生。欧洲自文艺复兴时期起出现了射影几何。18 世纪，G. 蒙日应用分析方法于形的研究，开微分几何的先河。C. F. 高斯的曲面论与 B. 黎曼的流形理论开创了脱离周围空间以形作为独立对象的研究方法。19 世纪 F. 克莱因以群的观点对几何学进行统一处理。此外，如 G. 康托尔的点集理论扩大了形的范围，H. 庞加莱，又创立了拓扑学，使形的连续性成为几何研究的对象。这些都使几何学面目一新。

在现实世界中，数与形，如影之随形、难以分割。中国的古代数学反映了这一客观实际，数与形从来就是相辅相成，并行发展的。例如勾股测量提出了开平方的要求，而开平、立方的方法又奠基于几何图形的考虑。二次、三次方程的产生，也大都来自几何与实际问题。至宋元时代由于天元与相当于多项式概念的引入，出现了几何代数化。在天文与地理中的星表与地图的绘制，已用数来表示地点，不过并未发展到坐标几何的地步。在欧洲，14 世纪 N. 奥尔斯姆的著作中已有关于经纬度与函数图形表示的萌芽，而 17 世纪 R. 笛卡儿提出了系统的把几何事物用代数表示的方法及其应用，在其启迪之下，经 G. W. 莱布尼茨、I. 牛顿等的工作，发展成了现代形式的坐标制解析几何，使数与形的统一更臻完美，不仅改变了几何证题过去遵循欧氏几何的老方法，还引起了导数的产生，成为微

积分产生的根源。这是数学史上的一件大事。在这一世纪中，由于科学与技术上的要求促使数学家们研究运动与变化，包括量的变化与形的变换（如投影），还产生了函数概念和无穷小分析即现在的微积分，使数学从此进入了一个研究变量的新时代。18 世纪以来，以解析几何与微积分这两个有力工具的创立为契机，数学以空前的规模迅猛发展，出现了无数分支。由于自然界的客观规律大多以微分方程的形式表现，因而微分方程的研究，一开始就受到重视。微分几何基本上与微积分同时诞生，高斯与黎曼的工作又产生了内在的现代微分几何。19、20 世纪之交，庞加莱创立了拓扑学，开辟了对连续现象进行定性与整体研究的途径。对客观世界中随机现象的分析，产生了概率论。第二次世界大战军事上的需要以及大工业与管理的复杂化产生了运筹学、系统论、信息论、控制理论与数理统计学等学科。实际问题要求具体的数值解答，产生了计算数学。选择最优途径的要求又产生了各种优化的理论、方法。力学、物理学同数学的发展始终是互相影响互相促进的，特别是相对论与量子力学推动了微分几何与泛函分析的成长。此外，在 19 世纪还只用到一次方程的化学和几乎与数学无缘的生物学，都已要用到最前沿的一些高深数学。19 世纪后期，出现了集合论，还进入了一个批判性的时代，由此推动了数理逻辑的形成与发展。也产生了把数学看作一个整体的各种思潮和数学基础学派。特别是 1900 年 D. 希尔伯特关于当代数学重要问题的演讲，以及 30 年代开拓以结构造概念统观数学的法国布尔巴基学派的兴起，对 20 世纪数学发展的影响至深且巨。科学的数学化一语也往往为人们所乐道。数学的外围向自然科学、工程技术甚至社会科学不断渗透扩大并从中吸取营养，出现了一些边缘数学。数学本身的内部需要也孳生了不少新的理论与分支。同时其核心部分也在不断巩固提高并有时作适当调整以

适应外部需要。总之，数学这棵大树茁壮成长，既枝叶繁茂又根深蒂固。本卷对于数学各分支与各种流派有较全面的介绍。

在数学的蓬勃发展过程中，数与形的概念不断扩大，日趋抽象化，以至于不再有任何原始计数与简单图形的踪影。虽然如此，在新的数学分支中仍有着一些对象和运算关系借助于几何术语来表示，如把函数看成是某种空间的一个点之类。这种做法之所以行之有效，归根结底还是因为数学家们已经娴熟了那种简易的数学运算与图形关系，而后者又有着长期深厚的现实基础。而且，即使是最原始的数字如1、2、3、4，以及几何形象如点与直线，也已经是经过人们高度抽象化了的概念。因此，如果把数与形作为广义的抽象概念来理解，则前面说到的把数学作为研究数与形的科学这一定义，对于现阶段的近代数学，也是适用的。

由于数学研究对象的数量关系与空间形式都来自现实世界，因而数学尽管在形式上具有高度的抽象性，而实质上总是扎根于现实世界。生活实践与技术需要始终是数学的真正源泉。反过来，数学对改造世界的实践又起着重要的、关键的作用。理论上的丰富提高与应用的广泛深入在数学史上始终相伴相生，相互促进。但由于各民族各地区的客观条件不同，数学的具体发展过程是有差异的。大体说来，古代中华民族以竹为筹，以筹运算，自然地导致十进位值制的产生。计算方法的优越有助于对实际问题的具体解决。由此发展起来的数学形成了一个以构造性、计算性、程序化与机械化为其特色，以从问题出发从而解决问题为主要目标的独特体系。而在古希腊则着重思维，追求对宇宙的了解。由此发展成以抽象了的数学概念与性质及其相互间的逻辑依存关系为研究对象的公理化演绎体系。

中国的数学体系在宋元时期达到高峰以后，陷于停顿且几

至消失。而在欧洲，经过文艺复兴、宗教革命、资产阶级革命等一系列的变革，导致了工业革命与技术革命。机器的使用，不论中外都由来以久。但在中国，则由于明初被帝王斥为奇技淫巧而受阻抑。在欧洲，则由于工商业的发展与航海的刺激而得到发展。机器使人们从繁重的体力劳动中解放出来，又引导到理论力学和一般的运动和变化的科学研究。当时的数学家们积极参与了这些变革以及相应数学问题的解决，产生了积极的效果。解析几何与微积分的诞生，成为数学发展的一个转折点。17 世纪以来数学的飞跃，大体上可以看成是这些成果的延续与发展。

20 世纪出现各种崭新的技术，产生了新的技术革命。特别是计算机的出现，使数学又面临一个新时代。这一时代的特点之一将是部分脑力劳动的逐步机械化。与 17 世纪以来数学之以围绕连续，极限等概念为主导思想与方法不同，由于计算机研制与应用的需要，离散数学与组合数学开始受到重视。计算机对数学的作用已不限于数值计算，符号运算的重要性日趋明显（包括机器证明等数学研究）。计算机还广泛应用于科学实验。为了与计算机更好配合，数学对于构造性、计算性、程序化与机械化的要求也显得颇为突出。代数几何是一门高度抽象化的数学，最近出现的计算性代数几何与构造性代数几何的提法，即其端倪之一。总之，数学正随着新的技术革命而不断发展。

数学概况及其发展*

数学，这门基础学科，已经越来越渗透到各个领域，成为各种科学技术、生产建设、以至日常生活所不可缺少的有力武器。在现代的科学技术中，如果不借助数学，不与数学发生关系，就不可能达到应有的精确度与可靠性。就科学来说，数学又是通向一切科学大门的钥匙，不仅所谓精确科学，如物理学、化学等已越来越需要较深较多的数学，甚至过去认为以描述为主，与数学关系不大的生物学、经济学等，也处于日益“数学化”的过程之中。这正象马克思早就指出过的那样，“一种科学只有在成功地运用数学时，才算达到了真正完善的地步”。

数学研究的对象是现实世界中的数量关系与空间形式。数与形，这两个基本概念是整个数学的两大柱石。整个数学就是围绕着这两个概念的提炼、演变与发展而发展着的。数学在各个领域中千变万化的应用也是通过这两个概念而进行的。社会的不断发展，生产的不断提高，为数学提供了无穷源泉与新颖课题，促使数与形的概念不断深化，由此推动了数学的不断前进，在数学中形成了形形色色、多种多样的分支学科。这不仅使数学这一学科日益壮大，蔚为大成，而且使数学的应用也越来越广泛与深入了。

我们将以数与形这两个概念为中心对数学的概貌先作一简

---

* 本文摘自《现代科学技术简介》，科学出版社，1978年。

单描述。

一、数学是研究数与形的科学

大体说来，数学中研究数量关系或数的部分属于代数学的范畴。研究空间形式或形的部分，属于几何学的范畴。此外，数与形是有机联系而不是相互割裂的。远古时代，关于长度、面积、体积的量度，我国宋元时代出现的几何代数化，以及17世纪的解析几何，把形与数这两个概念沟通了起来（因而也把几何与代数这两者沟通了起来）。近代函数概念与微积分方法的出现，在数学中形成了系统研究形、数关系的分析学，成为近代数学中发展最迅速的部分。几何、代数、分析三大类数学，构成了整个数学的本体与核心。在这一核心周围，由于数学通过数与形这两个概念与其它领域的互相渗透而出现了许多边缘学科与交叉学科。这是整个数学王国的一个总的轮廓。

先从数说起

最简单最基本的也是从远古时起人类就不得不与之打交道的数，乃是正整数或自然数：

$$1,\ 2,\ 3,\ 4,\ 5,\ \cdots$$

在正整数之间有两种最简单的运算：加法与乘法。研究整数之间的联系与规律的学问叫做数论。从乘法产生了素数* 的概念，例如6（$=2\times3$）是非素数，而7由于不能分解成两个比7更小的正整数的乘积而是素数（1不算素数）。正整数的一个基本性质是，它总可以表示成若干个素数的乘积，例如12＝

---

* 对于大于1的整数来说，若除它本身与1之外再没有其它因子，则称此数为素数。

$2^2 \times 3$，$18 = 2 \times 3^2$，$45 = 3^2 \times 5$ 等，而且这种表示方法只有一种。素数

$$2，3，5，7，11，13，17，19，23，\cdots$$

在整个正整数序列中的分布是极不规则的，这个素数分布规律的探求产生了许多迄今没有解决的著名难题，哥德巴赫（Goldbach）问题* 就是其中之一。这些难题反映了加法与乘法之间的矛盾，用初等方法对这些问题是无能为力的。微积分发明以后，数学家们开始用所谓解析方法来研究数论，开创了解析数论这一学科。我国在哥德巴赫问题上的第一流成果，就是用了解析方法而获得的。

19 世纪中，数学家把整数概念大大扩大了。例如，我们可以考虑所有形如 $a + b\sqrt{2}$的数，其中 a 和 b 则是通常的整数（正、负或零），称这些为$\sqrt{2}$域中的“整数”。它们也可以相加、相乘，因之也可以定义“素数”。可以证明任一$\sqrt{2}$域中的“整数”基本上只有一种方法把它表示成若干个“素数”的乘积。但如果考虑所有$\sqrt{-5}$域中的“整数”，即形如 $a + b\sqrt{-5}$的数，a 和 b 仍是通常的整数，情形就大不相同了。例如，21 与 9 就都有两种完全不同的方法表示成“素数”的乘积：

$$21 = 3 \cdot 7 = (1 + 2\sqrt{-5})(1 - 2\sqrt{-5}),$$
$$9 = 3^2 = (2 + \sqrt{-5})(2 - \sqrt{-5})。$$

数学家为了要克服这一困难，创立了理想数或简称“理想”的理论，$\sqrt{2}$域与$\sqrt{-5}$域也推广到了一般的代数数域，这种域上

---

* 哥德巴赫提出一个论断：“任何一个大于 4 的偶数均可表示为两个素数之和”，即“1＋1”。我国数学家陈景润同志已证明了：“大偶数均可表示为一个素数加不超过两个素数的乘积”，即“1＋2”，把这个论断的证明向前大大推进了一步。

整数的理论已发展成为一个当前很活跃的数学分支，叫代数数论，“理想”也已成为现代抽象代数学中最基本的概念之一。

数的概念是不断发展的，从整数出发，人们逐步引进了分数、小数、正负数、无理数等概念而形成了实数系统。由于解代数方程的需要，又引入了虚数、复数而构成了复数系统。这些实数（或复数）之间可以加、减、乘、除，且这些运算遵守通常所谓交换、结合、分配等等规律。数学家们把具有这些运算并满足这种规律的实数或复数全体，称为实数域或复数域。

随着数学的发展，人们又引进了与通常的数很不相同的量，但却具有与数相类似的运算。例如，在力学中力可表示成一个向量，两个力 F_1、F_2 的合力是 F_3 时，可以记作 $F_1+F_2=F_3$，而这种加法也遵守交换律与结合律。又如绕固定点 O 各作旋转 q_1、q_2，如果先作旋转 q_1，再作 q_2 所得是一绕 O 的旋转 q_3，而先作 q_2，次作 q_1 所得是绕点 O 的旋转 q_4，就记作 $q_1q_2=q_3$ 与 $q_2q_1=q_4$。一般说来，q_3 与 q_4，是不同的旋转。19 世纪时英国的数学力学家汉密尔顿把绕点 O 的旋转视作所谓“四元数”。在四元数间也可以相加、相乘，但其乘法不遵守交换律，即 $q_1q_2\neq q_2q_1$。

正象旋转之被视为四元“数”那样，许多在数学中陆续出现带有某种运算的事物，如向量、张量、矩阵以至更抽象的元素，都不妨视之为某种广义的“数”。这些“数”都以可以“运算”为其特征。同时，数学家也把研究重点逐渐从“数”的本身性质转移到“数”与“数”间的运算上面。带有某种运算的“数”的集体统称为代数系统。依据运算规律的不同而有各种不同的代数系统，并具有种种各别的名称，例如群、环、域以及环上的模，与域上的代数，等等。

由于群、环、代数等代数系统在数学中的广泛出现，又由于各种理论与应用中出现的问题，最后往往归结为某种代数系

统的研究，代数系统的一般理论发展成了分支繁多（如群论、环论等）的代数类数学，或所谓近代抽象代数学，它已成为整个数学最基本的工具之一。

再谈谈形

空间或几何形态是物质存在的躯体与外壳，人类首先注意到的物体的几何形态是大、小、方、圆，诸如长度、面积、体积、相似性等等，它们由于生产上的直接需要而首先从丰富的实践经验总结上升成为理论。在古代，我国与希腊形成了都以度量性为主但各有内容特色的不同几何体系。文艺复兴时期，绘画与建筑的实践经验，以及拿破仑时代军事上工程作图的需要，图形平直透视一类性质的研究，促成了另一种新的所谓投影几何学的出现。它的研究几乎贯串了整个 19 世纪。到了上一世纪晚期，另一类几何性质——空间的连续性与连通性——开始受到了重视，由于这些性质虽基本而隐晦，因而不易被发现与处理，只是由于科学的不断发展，许多数学问题都导致空间这类性质的研究，才在较近时期，即 19、20 世纪之交，形成了一门崭新的几何学分支——连续几何学或拓扑学。它的蓬勃发展乃是本世纪数学的一个特色。

几何研究的对象与方法也有很大的变化。例如，以光滑曲面为对象，通过引入弧线长度概念而建立了微分几何学。以后又推广到高维的光滑流形，并由于拓扑学的发展而开展了流形全局性或整体性的几何拓扑研究，引进了各种示性类与示性数。这些类与数已在最近被应用于磁单极与“基本”粒子等物理学的基本理论研究。又如，由于求解一般的多项式方程组，开展了由这种方程组的解答所构成的空间，即所谓代数簇的研究，形成了所谓代数几何学这一分支。解析几何的出发点是，引进了坐标来表示点的位置。同样，对任一代数簇也可以引进

坐标，这为代数几何的研究提供了一个有力的工具。除了研究光滑流形与代数簇这种特殊类型的空间（由于它们的特性得以应用发展较成熟的分析，即微积分方法与代数方法）的几何学以外，数学家又考虑了最一般的点的集体所构成的空间，研究它们的连续性质与度量性质，形成了所谓点集拓扑学与测度论这些分支学科。

数与形的联系

17 世纪是数学发展历史上一个划时代的新阶段的开始。这一时期，创立了解析几何，又出现了变量与函数的概念，把数学中的两大基本概念形与数紧密地联系在一起。所谓函数，即是定义在某些空间上的数量的分布。例如，在大气层中的压力、温度等等物理量的分布，即是定义在大气层空间上的压力函数、温度函数等等。17 世纪通过用微分表达变化，与用积分表达积累，又创立了研究函数的变化与积累的微积分方法，使数学得到了一个认识自然的有力武器，面目为之一新。自然界的规律往往表现为某些物理量之间的变化与积累的相互制约关系，在数学形式上，则表现为定义在某些空间上的函数间的微分积分方程。举例来说，所谓气象预报，无非是根据过去一段时期，对各地压力、温度、降雨量等等的实测数据，以及表达气象变化规律的这些函数间的微分方程，用数学方法推算出今后一段时期内的这些函数数值，以预报气象特征而已。这种帮助认识自然，进而改造自然的普遍而有力的数学方法，使相应的一些数学分支，如函数论、微分方程、数学分析等成为三百多年来数学发展的主流，构成了庞大的分析一类数学，并由于要解决有关问题，而促使一些新的数学分支，如微分几何学、拓扑学、泛函分析、计算数学等的出现与迅速成长。

形与数这两者并不是互相割裂的，早在产生数学的萌芽时

期，就通过长度、面积与体积的量度而把形与数联系了起来。我国宋元时期更系统地引进了几何代数化的方法，把一些几何特征用代数式来表达，几何关系则表达为代数式间的代数关系，成为解析几何的先驱，使空间形式的研究归结为较成熟也容易驾驭得多的数量关系的研究。在近代的数学中，这个方法原则也一直在使用着。例如，在拓扑学中，通过引进一些数（如贝蒂数）或代数系统（如同调群，同伦群等）来表达拓扑空间的连续性与连通性，然后用代数方法对这些数与代数系统进行分析而获得拓扑空间几何性质方面的信息。依据这种思想，在19世纪末开始建立起来的代数拓扑学，成为拓扑学中最有活力的分支，在本世纪中有着极大的发展，对整个数学也有不小的影响。

不仅几何学由于代数化而获得了有力的武器，而且代数学（以及分析学）也往往由于借用了几何术语，运用几何类比而得到了新的生命力，促进了它们的发展。例如，早在18世纪中，法国数学家拉格朗日就把时间因素作为与三个空间坐标并列的第四个坐标而引入了四维空间，推动了力学的研究。同样，力学家与物理学家往往把各种物理参数作为不同坐标而引进了高维的相空间等概念，使几何方法得以在物理学中发挥作用。现代的相对论，即在这种方法下与四维时空流形的几何学研究不可分离。现代数学中还有一个常用的方法，即把一个个函数看作一个个“点”，而把某类函数的全体看作一个“空间”，函数间的相异程度看作“点”之间的“距离”，由此得到了各种无穷维的函数空间。一个微分积分方程组的求解，往往归结为求相应函数空间中一个几何变换的不动点问题。这样，不仅分析的问题具有了几何“直观”的意义，而且可以运用近代几何拓扑，以至抽象代数学的有力方法。由此在分析类数学中产生了泛函分析这一活跃的分支，在现代自然科学甚至工程

技术的应用中起着极其重要的作用。

数学发展中的边缘学科

几何、代数与分析等类的学科构成了数学的本体与核心。随着数学本身的发展，以及科学技术与生产实践对数学的新的要求，在这一核心的周围又形成了许多外围学科与边缘学科。例如，客观世界中大量存在的随机现象的研究自 18 世纪以至近代逐渐形成了概率论、随机过程论与数理统计等随机类学科。近几十年来，由于现代生产与国防建设的需要，对资源、设备和条件的合理使用与统筹规划，形成了优选学、规划论、对策论、排队论等运筹类学科。现代大工业要求对工程系统的操作能更可靠与更经济，并能自动控制，特别是由于航天技术等，对控制系统高精度的要求，出现了一门介于数学和工程之间的边缘学科控制论。工程乃至生物中各种复杂的信息传递的研究，推动了边缘学科信息论的诞生和发展。大量应用数学问题的解决，最终需要取得有一定精确度的数据而出现了各种具体的计算理论与计算方法，由此促进了计算数学的发展并产生了一门数学——计算机科学的边缘学科。数学与各种科学的相互渗透又出现了物理数学、生物数学与经济数学等交叉学科。此外，数学各基础学科间的相互渗透也在数学核心内部产生了各种综合性学科。例如，流形上几何、拓扑与微分方程的综合研究，形成了一门新的分支“大范围分析”，随机函数的微分方程论研究，又产生了概率论与分析学的综合性学科“随机微分方程论。”象这一类的综合性学科正在不断涌现。

另一面，17 世纪以来由于解析几何与微积分这种强有力的新工具的出现，数学家们忙于应用这些工具解决科学与技术中的一大堆问题，并为新方法的成功所陶醉，对于所依据的理论是否可靠，基础是否扎实，未遑遐顾。到了 19 世纪，数学

家们已越来越感到谬误与正确杂陈的局面之无法容忍，许多概念必须澄清，数学也必须置于严密基础之上。在这种形势下，从 19 世纪中叶以来，主要在一些德国数学家的倡导之下，对数学进行了一场批判性的检查运动。这场运动不仅使数学奠定了严实的基础，并产生了公理化方法，以及一些集合论、实变函数论、点集拓扑学、抽象代数学等新颖学科。特别是，数学推理本身的分析与形式化产生了一门影响巨大的学科“数理逻辑”。

此外，数学史料的征集与整理也被重视起来，出现了一些篇幅巨大的数学史著作。近年来，数学史已进入了对数学思想与方法的历史演变和分析批判的研究阶段。对中国古代的数学过去或付阙如、或多歪曲的情况，也已受到了重视，开始走上了正确理解与分析的时期。我国的数学工作者们，应该责无旁贷，把对我国数学史的研究重任担当起来。

这样，以几何、代数、分析等类学科为核心，以及围绕此核心周围的形形色色的许多学科错综复杂地交织在一起成为数学王国的一幅宏大而绚丽的蓝图。

数学学科的盛衰兴替

当然，学科并不是一成不变的。随着时间的推移，新的学科不断产生，旧的学科却有时不免销声匿迹默默无闻。绵亘于整个 19 世纪的投影几何学，由于基本上已搞清楚而被作为档案搁置在图书馆的书架上。在本世纪 20 年代盛极一时的射影微分几何学，则由于后来发现意义不如预期而受到了冷落。在 19 世纪对代数与几何都极重要的不变式论这一学科，则其盛衰兴替已有几起几落。某些学科，如代数几何学，其面貌又经常在改变之中，诸家学说纷纭，从未获得定型。另一面，某些 18、19 世纪，甚至更悠久且已消逝的理论，却又重新被发掘

出来，在新的观点、新的方法之下，成为现代数学中很活跃的研究课题。例如，一门作为微积分最早应用的古老学科“变分法”，本来消沉已久，近一、二十年来，却以另一种控制论的新面目，出现在技术领域中显得十分活跃。这两三百年来的不少经典著作，也被重新翻印和钻研之中。

然而，不管数学各个学科经历着怎样的分、合、改、变，也不管数学内部如何奔腾澎湃，数学王国的疆土总是在不断扩张之中，而且始终是由形与数两大基本概念所统治之下的“世袭”领土。伟大的革命导师恩格斯对数学所作出的“纯数学的对象是现实世界的空间形式和数量关系”的这一精辟刻划不仅是对数学过去的总结，也是对数学现状的忠实描绘，又是对数学未来发展的指导与准绳。

二、数学发展的未来

展望未来，数学科学丰富多彩的广阔天地之日新月异，似乎使数学不得不分成许多日益庞杂的分支，各分支间的联系，也有日益削弱的趋势。许多专门家们往往只能偏处一隅之地，彼此很难相互了解。然而这只是数学发展过程中的现象，也只是事物的一个方面，随着数学的不断前进，也必然会不断出现更锐利的工具与更简单的方法，使数学不断更新，删繁就简，去粗存精。在历史上，原来只能用天才式的艰难推理才能从事的初等几何学，通过我国宋元时代的几何代数化，与 17 世纪时的解析几何，已变得平淡无奇。从有史以来直到 17 世纪包括解析几何与微积分为止的全部数学，不论岁月如何漫长，卷帙如何浩繁，内容如何丰富，基本上都已（或应已）压缩在十来年的中小学教科书中。这种现象也正在重复。

为了要使庞大的数学知识变得简而且精，数学家们经常依据数学各领域间潜在的共性，提出统一数学各部分的新观点新

方法来。例如，在 19 世纪后期，德国厄兰格的数学家克莱因提出用“群”的观点来统一当时杂乱的各种几何学的方案，迄今称为厄兰格计划。本世纪 20 年代，美国伯克霍夫又提出了“格”的概念，以统一代数系统的各种理论与方法。上一世纪与本世纪之交，出现了公理化运动，以公理系统作为数学统一的基础。本世纪 30 年代，法国的一个数学学派布尔巴基，除了继承了公理化运动以外，又提出了结构概念，把数学的核心部分在这概念之下统一成一个整体，而依研究结构的异同来划分成不同领域。与此大约同时，美国麦克莱恩与艾伦伯格又提出范畴与函子理论，数学的分门别类即以研究所属范畴为依据，以此作为统一数学的基础。总之，各种不同学派，根据他们对数学的不同认识，提出了多种多样统一数学，以建立数学体系的不同方案来，其终极目的则大体上是相同的。或许将来会出现一种新观点、新方法、新理论，把目前的数学统一起来。但是，这种统一不是绝对的、静止不变的，而是暂时的、相对的，对立的斗争则是绝对的。随着科学技术的不断发展和深入，以及各学科之间的相互渗透，将对数学提出大量的千差万异的新课题，使数学研究得更深入。这些问题的解决，反过来又促使数学更进一步发展。

对于数学未来发展具有决定性影响的一个不可估量的方面是计算机对数学带来的冲击。在不久的将来，电子计算机之于数学家，势将与显微镜之于生物学家，望远镜之于天文学家那样不可或缺。现在的计算机通过小型化而成为每个数学家的“囊中之物”，这一设想势将成为现实，数学家们对这些前景必须有着足够的思想准备。

除了一些人所共知的作用外，计算机还提供了一个有力工具使数学有可能象其它自然科学一样，跻身于科学实验的行列。19 世纪的数学家高斯，为了要发现整数性质的规律性，

首先对各种特殊情况做大量艰苦的计算工作以为试探。高斯关于整数论的一些著名定理，用他自己的话来说，就是通过这种“系统尝试法”而发现的。现在这种手工业式的系统尝试可以用电子计算机来代替了。美国数学家乌拉，为了要探讨应用中广泛出现，而现代数学又显得还无能为力的非线性现象，在电子计算机上进行试验，发现了一些规律，近几年来，在应用数学上的一个重大突破（所谓微分方程孤立子解的获得），就是首先在计算机的荧光屏上发现的。

计算机对数学的另一个重大作用乃是对数学研究作为脑力劳动在方式上带来的革新。数学，不论是学习还是创新，最耗时费力的劳动，往往消耗在定理的证明上，而不是在真理的发明、发现上。事实上，一个定理即使对其证明在逻辑上经历了严格细致的逐步检验也无非是说明知道定理正确无误而已，还不足说明真正“懂得”了这个定理。自然，证明是完全必要的，证明的严密性也是完全必需的，但更重要的应是定理之为何发明，如何发明，起何作用等这一类问题。电子计算机已使人们从烦琐复杂而又十分单调的加减乘除的劳动中解放出来。某些数学定理的证明，完全可以借助于电子计算机来完成。事实上，某些或某类定理的证明，可以避免通常虽简美但奥秘因而颇为艰难的方式，而采用虽烦琐但刻板因而较为容易的方式。换言之，即使质的困难转化成量的复杂，而后者对于电子计算机来说是轻而易举的，因而得以使定理证明化难为易。这样电子计算机就可以使人们从某些逻辑推理的脑力劳动中解放出来。因而使数学家得以把聪明才智更多地用到真正创造性的工作上去，这是当前数学发展中值得也是应该认真考虑的问题。

我国的数学有着悠久的历史与优良的传统，本来位于世界的先列，只是近几百年来才落后了。

解放以后，我国数学的发展曾呈现一片兴旺景象，出现了与国际先进水平的差距越来越小的势头。但近几年来，由于“四人帮”的严重干扰和破坏，使我国数学的发展大大落后于形势的需要。

在党中央向科学技术现代化进军的伟大号召下，我国应急起直追，不仅赶而且超，在本世纪余下的 23 年内，我国的数学工作者，不仅应在数学上有所创造，有所发明，有所前进，以丰富数学的内容，拓展数学的领土，增强数学的基础研究，扩大数学的应用，而且应以马列主义、毛泽东思想为指导，以数学的渊博知识与深湛认识为基础，提出我们自己的观点与方法，建立几个具有我国特色的学派，涌现出一批在国际上有影响的数学家，为独立自主地解决社会主义建设中提出的数学问题，为四个现代化作出较大的贡献。这将是我国老、中、青数学家们所面临的一项十分艰巨而又十分光荣的历史任务。

关于教材的一点看法*

为了提高整个中华民族的科学文化水平，加速实现四个现代化的步伐，必须大幅度提高基础教育的质量。因此，把较高的基础知识有条件地适当地纳入较低的基础教材之内，已是一项提到教材改革日程上来的问题。不弃旧无以纳新，为了安排较新较高级的内容，某些陈旧的内容必须有所压缩，甚至从原来教材中淘汰出去，这也是无可非议的事。

这种教材上的弃旧纳新，在历史上并不是什么新鲜事。远的不谈，只就近几十年来的情况看，也可证明这一点。举几个例子：解放前的数学教材，初中一年级整整一年学的是算术，主要内容是鸡兔同笼一类的四则难题，解放后很早就一笔勾销，初中一年级就学代数。解放前的初等几何，往往在九点圆一类的难题，趣题上消耗不少精力，解放后也早已在课程中排除，让位于解析几何。这些改革，现在大家都承认是必要措施，很少人会提出异议了，但在当时决不是顺顺当当的。

究竟纳什么新弃什么旧，纳又纳到什么程度，弃又弃到什么程度，则应根据具体情况作具体分析，郑重从事。相信教育界和数学界的同志，本着对我国人民与我国科学事业的负责精神，会各抒所见，提出切实可行合于实际的良好意见来。在这里，我愿抛砖引玉，说点自己的意见。我个人认为，初等微积分应该处于最优先考虑的地位，它的意义作用比之所谓集合矩阵之类重要

---

* 本文摘自《光明日报》，1978 年 7 月 29 日。

得多，而且学起来对中学生并无多大困难。其次，还可考虑增加一些有关计算器使用的项目，这在国外的教学改革中，也已经成为不少中学新教材的必修内容了。

上面的意见，是对于较长期的改革来说的。在短期内，我们也不能不顾及某些现实情况。由于“四人帮”对教育的摧残和破坏，这些年来中学数学水平大大降低。对于正在中学就学的同学们来说，当务之急，还在于加强计算能力与逻辑推理等的基础训练，努力提高数学质量。宁可少些，但要好些。自然，长期与近期措施之间，并不是截然分开，而是可以并行不悖的。

数学教材一方面应有弃旧纳新的准备，另一方面也应注意必要的相对稳定性。科学发展日新月异，数学的创新也层出不穷。但教材毕竟与科学的创新不同。如果一味求新，而那些新的内容是不是能经得起时间的考验，往往很难预测。改得不好，会造成灾难。我们只能将已经历过较长时间的考验，并肯定有广泛应用前途的内容纳入新教材，而不能凭主观冒失从事，违背最起码的认识规律。例如，我们不能让学生在没有经过具体数字运算的基本训练的情况下，就来学习以文字代替数字的初等代数或使用计算器。同样，不能在还没有掌握直线与圆的基础知识之前，就学习几何的代数化与解析几何。这些道理，本来不言而明，但国外某些数学教学改革却曾因违背这些常理而招致灾难性的后果，不能不引以为戒。

数学与四个现代化*

敬爱的周总理曾经指出：实现四个现代化的关键在于实现科学技术现代化。而科学技术的现代化，处处离不开数学。数学可以说是基础科学中的基础学科，也可以说是科学现代化的基础。

革命导师马克思和恩格斯都十分重视数学。马克思指出：一种科学只有在成功地运用数学时，才算达到了真正完善的地步。他自己还写出了精辟的微积分的论文。恩格斯在他写的《自然辩证法》和《反杜林论》中，对数学产生的历史，发展的动力，以及应用与作用，都有非常精辟的论述。他还给数学下了一个最恰当的最有概括性的定义。他说："纯数学的对象是现实世界的空间形式和数量的关系，所以是非常现实的材料。"这个定义不但总结了数学的过去，描述了数学的现状，还指出了数学今后的发展方向。

恩格斯下的定义，指出了数和形是数学中的两大基本概念。事实上，整个数学就是围绕这两大基本概念的演变而发展的，同时也通过这两大基本概念应用到各个不同的领域中去。例如造房子，要进行各种力学和材料的计算，这就离不了数；还要画出设计图纸，这就离不了形。所以在各种科学研究、生产建设和日常生活中，数学都是不可缺少的武器。

---

* 本文摘自《科学家谈数理化》，中国少年儿童出版社，1978年。

在小学和中学的数学课本里，我们接触到了整数、小数、分数、正数和负数、实数和虚数，以及它们之间的各种运算，如加、减、乘、除、开方等等，这些构成了以数量为主的算术和代数部门；我们又接触到了点、线、方、圆之类的图形，以及它们的性质和相互关系，如面积、体积、全等、相似等等，这些构成了以空间形式为主的几何部分。我们还要接触到表示几个数量之间相互制约的关系，就是函数概念，以及它们的图形所表示的变量、极限和坐标等等，这就进入了把数和形这两大基本概念沟通起来的解析几何和初步微积分的领域。我们还要接触到这些基础知识在各种实际运用中的具体问题。上面讲的，就是小学和中学数学课本的大致轮廓。

小学和中学数学课本的全部内容，都是人类在漫长的岁月中积累下来的数学知识的精华。在探索这些知识时，不少人付出了艰苦的劳动，后来一再加工提炼，才使这些知识变得简单易懂。譬如说负数的概念，现在大家都很容易理解。但是在 16 世纪，欧洲最博学的数学家也感到难以理解，甚至把负数叫做“虚构的数”，“荒谬的数。”直到 18 世纪，英国的一些大学还反对使用负数。而在我国，古代的数学是十分先进的。有一部数学经典著作，叫《九章算术》，是在两千年以前西汉初年写成的。里面有一章《方程》，记载了 18 个联立方程组的问题，未知数从两个到 5 个，在解这些问题的时候已经运用了正数和负数。我国古代的这个辉煌成就，说明了我国人民不但有高度的分析问题和解决问题的能力，还有高度的抽象能力，善于以广泛的实践为基础，提炼出有普遍意义的概念。

我国古代在数学方面的许多成就，经历了长期的历史考验，直到今天还有现实意义。例如在多种多样的实际应用中要解决的联立方程组，未知数不是 3 个 5 个，而是几十个几百个。解这样繁复的联立方程组，用笔算是无论如何办不到的。在电子计

算机出现后不久，一位专家曾经估计：一个有 26 个未知数的联立方程组，在电子计算机上用 18 世纪在欧洲出现的“克莱默方法”来解，得花十万万万万年；如果用消去法来解，只要用 3 秒钟就够了。这种消去法，我们在中学的代数课本上就可以学到，也是我国两千年前创造的方法，也载在《九章算术》上。

我国古代在数学方面，创造了丰富多采的理论和方法，为人类作出了很大的贡献。但是由于长时期的封建统治和近一百多年来的帝国主义侵略，生产得不到发展，数学的发展也停滞了。而欧洲，由于文艺复兴和工业革命，资本主义制度取代了封建制度，解放了生产力，打开了数学发展的新局面，出现了解析几何、微积分以及内容丰富的现代数学。这些现代数学知识，都是我们实现四个现代化所必不可少的。举例来说，搞工业总要遇到如何能多快好省的问题。最简单的这一类问题可以归结成这样的形式：求一个函数的极大值或极小值。有了微积分的知识，这一类问题的解决就变得轻而易举了。现代的大工业中，还出现了许多新的求极大值和极小值的问题：有的函数关系根本不知道，有的虽然知道了函数关系，但是变数受着一些不等式的限制。解决这样的问题，微积分也无能为力，数学家们就创造了一些新的理论和方法，解决前一类问题的有优选法，解决后一类问题的有规划论。

不但工业要用到数学，农业机械化和国防现代化也都要用到高深的数学。农业现代化首先要实现机械化，在农业机械的制造中，就有大量的数学问题。例如拖拉机上一个小小的犁铧，犁面做成什么形状，翻土的作用才能起得好，用到的数学就不简单。国防现代化要设计新武器，要提高作战能力，如导弹形体的设计，要考虑到振动不致影响方向的问题。在指挥空战中，要考虑到敌机如何逃避和我机如何追逐的所谓追逃问题。所有的问题都要用高深的数学知识来解决。科学技术的现代化，遇到的

数学问题就更多了。毛主席有过重要指示，要进行基本粒子的研究。这是目前世界上最活跃的一门科学，几乎所有的数学工具都运用上了。数学作为一种应用的武器，可以说无孔不入。四个现代化如果离开了数学，将寸步难行。

消除对数学的神秘感*

——推荐《数学译林》

一个人从小学、中学到大学，在各项课程中，投入时间最多的，无过于数学。一个人涉猎群书，在各种读物中，最感到费解的，也无过于数学。数学的各种书刊，不仅对于非数学家是一种禁区，即使是数学家，从事数学这一领域研究的专家要阅读另一领域的著作，对于其中出现的许多概念符号（更不用说推理论证），也往往会感到玄妙莫测，必须花费很大的气力才能得到一个大致的印象与了解。

如何填平数学与非数学之间的鸿沟，消除非数学家对数学的神秘莫测感？如何消除数学各个不同领域专家之间的隔阂，使他们不致隔行如隔山，以增进彼此间的了解并进而交流合作？这都是值得重视的问题。

《数学译林》这一刊物，不同于一般的数学刊物，它多少有助于弥补上面所说的那一些缺陷。

《数学译林》是以译文为主的综合性数学刊物，其目的在于介绍国外数学的进展，普及现代数学知识，由此促进我国数学事业的发展与水平的提高。

《数学译林》辟有各项专栏。综合报告与专题介绍专栏就对某一领域或某一专题的历史与现状作较全面较通俗的论述，使非本专业的人能迅速获得对它的了解，甚至被吸引到这一方面

---

* 本文摘自《光明日报》，1985年4月19日。

来参加工作。

数学史与人物传记专栏评介数学发展中出现的重要事件与著名学者的重大贡献，从历史名人到当代标新立异的数学家，从个别人物到影响巨大的大学派，巨细不捐。

数学争鸣栏介绍数学上的重大争论，例如对纯粹数学与应用数学以及数学的作用等的许多不同看法。英国某数学权威推崇“无用”的数学为至高无上的数学的著名论点，以及另一有名学者对它的反驳，就登载在同一期中。

此外，如数学小品、数学教育、数学竞赛等栏也都各有特色。

尤其值得一提的是，在文章末尾空白之处，往往摘引了一些有名人物的名言警句，足以发人深省。

自然，数学究竟与其他科学不完全一样，即使是一篇通俗著作，牵涉到一些专门的论述时也会使读者感到困难，但这决不是主流。不论是专业性的数学专家，或是只对数学抱有兴趣的门外汉，都有可能从刊物中找到一些能使他们满足的作品。总的说来，这一刊物大体上能做到雅俗共赏。我自己可以说既是数学中的雅士，也是数学上的俗客，不论从哪一角度，都是这一刊物的爱好者与经常阅读者。为此特将这一刊物推荐给广大读者。

对中国传统数学的再认识*

(一)

中国传统数学源远流长，贡献巨大，对数学的发展有不可磨灭的作用，但自宋元时期出现最后一次高峰后，明代以来即趋衰微，至明末已几成绝学。从明清以至今日，实质上已退出了数学舞台。历代中外算家，也曾尝试阐释幸存至今的中国经典算书，但一些抱着轻蔑成见妄加诋毁者姑置不论，即使出于善意克服古文字的阻碍而从事认真钻研的学者，也往往或拘泥于西方数学的先入之见，或着眼于以现代的数学方法与成就理解古人著作，以西释中，以今议古，致使面目全非，掩盖甚至歪曲了中国传统数学的真实面目。试举数例如下。

西方早在欧几里得的《几何原本》中就已出现了素数的概念并证明了素数个数无穷。19世纪以来又出现了复域中整数、素数与因子分解等概念，并用解析方法研究素数的分布。由此创立的代数数论、解析数论等领域，都已成为现代纯粹数学的核心部分。与之相反，我国传统数学中从来没有出现过素数与因子分解一类概念，妄论其他。但是，我们不能因此就认为中国古时没有数论。

欧几里得几何中平行线公设的独立性一直是西方数学家们所关注的问题，也是几何发展的一个重要动力，19世纪时终于

---

* 本文摘自《百科知识》，1980年第7、8期。

摆脱了这一公设，创立了非欧几何等领域。与之相反，我国传统数学中从来就没有平行线概念的明显痕迹。但我们不能因此就认为中国古时没有几何学。

16 世纪以来，西方出现了以符号代表数字的作法，到 19 世纪发展起遵守各种运算规律的不同代数系统。对于方程，则阐明了根与系数间的关系，证明了必有复数解的代数基本定理，建立了能否用系数根式表达式求解的 Galois 理论。与之相反，我国古时未曾出现过用文字代表数字以及讨论根的性质一类工作。但我们不能因此就认为我国古时没有代数学。

欧几里得《几何原本》建立了由定义、公理、定理、证明构成的演绎系统，成为近代数学推理论证的典范。与之相反，我国典籍中从未出现过这种演绎证明的方式。但是，我们不能因此就认为我国古代数学没有逻辑思维，也不用逻辑推理。

早在古希腊时，西方就论证了不能用两整数之比表达的所谓无理数的存在。19 世纪又明确了实数概念与复数概念。与之相反，我国古代数学从未考虑过无理数或实数这样的概念，更没有复数的痕迹。但是，我们不能因此就认为我国古代没有数系统甚至没有数学。

因此，要真正了解中国的传统数学，首先必须撇开西方数学的先入之见，直接依据目前我们所能掌握的我国固有数学原始资料，设法分析与复原我国古时所用的思维方式与方法，才有可能认识它的真实面目。这一尝试自然并非易事。首先是资料问题，我国古籍散失者多而幸存者少。我国最重要的经典算书是《九章算术》，按钱宝琮说这最迟是在公元 50 至 100 年间的汉代定型成书。较之稍早见之于《汉书》本来还有许商、杜宗等算术，但只存其名。5 世纪南北朝祖冲之的《缀术》是一部名著，唐时被列为《算经十书》之一，为国子监算学生所必读，后终因难学而废黜，至宋元时即已失传而不详其内容。宋元时期曾刊行过大

量算书，但杨辉的某些著作只剩残卷。朱世杰的《算学启蒙》在我国本土已经绝迹而于 19 世纪初复得之于朝鲜。在杨辉、李冶、朱世杰的著作中曾提到过的当时的许多算书，更久已荡然无存。虽然如此，尽管古籍存者不多，但它们还是提供了一个足够清晰的轮廓，足以使人们看出我国传统数学固有的体系、方法与内容，以及一条相当清晰的发展途径。近年来，国内外学者对此论者颇多，笔者将随意引用他们的研究成果，但文中所表述的各种观点，则概由笔者本人负其全责。

中国传统数学的体系

西方的欧几里得体系着重抽象概念与逻辑思维以及概念与概念之间的逻辑关系。与之相反，我国的传统数学则基本上是一种从实际问题出发，经过分析提高而提炼出一般的原理、原则与方法以最终达到解决一大类问题的体系。与欧几里得体系相适应，它有一个以定义、公理、定理、证明所构成的表达形式。同样，我国的传统数学与它的独特体系相适应，也有一种独特的表达方式。大体说来，中国数学的经典著作大都以依据不同方法或不同类型分成章节的问题集的形式出现。每一各别问题又都分成若干个条目。条目一是“问”，提出有具体数值的问题。条目二是“答”，给出这一问题的具体数值答案。条目三称为“术”，一般说来乃是解答与条目一同种类型问题的普遍方法，实际上就相当于现在计算机科学中的“算法”，但有时也相当于一个公式或一个定理。条目四是“注”。说明“术”的依据与理由，实质上相当于一种证明。宋元以来，可能是由于印刷术的发达，往往加上条目五“草”，记述依据“术”得出答案的详细计算过程。

这里应该特别提出条目三“术”的作用。虽然条目一、二中的问与答都以具体数值表达，有时甚至术文本身也是如此，但不难看出所有术文都具有普遍意义。术文中即使带有具体数值，

这些数值并不起重要作用。如果以其他同类型的数值来代替，术文也依然行之有效。条目四的“注”或即证明也是如此。论证的正确性完全不依赖于原设数值的特殊性。例如，《九章算术》第九章勾股的第一、二、三的三个问题，都是以勾三、股四、弦五为例，知其二求其第三者。求法名为勾股术，术文说：“句、股各自乘，并而开方除之，即弦。”显然，这是从勾股求弦的一般方法，与具体数值三、四、五无关。勾股术的注或即证明也是如此。因此，问、答甚或术文中的具体数值只起着一种举例说明作用，同时也指出了术即一般方法的来历与动机。

西方古代数学中也并非没有以问题集形式出现的著作。值得一提的是，推测出现于 3 世纪的《丢番图集》一书。这本书也是以问题集的形式出现的，分成 6 卷，其中搜集了 189 个关于整数的问题。它是西方近代数论的源泉。所谓费马大问题，即是因 17 世纪时费马在该书边缘上写下了一个注解，但有关问题迄今未能解决而得名。丢番图一书不含有任何一般的方法。它所包罗的 189 个问题中所出现的特殊数字是有特殊作用的，每一个问题要用只适用于它自身特殊数字的特殊而往往又很奇特的办法，来求得其答案。因此，19 世纪德国著名的数学史家汉克尔(H. Hankel)在他的名著《上古与中古时期数学史》一书中曾经这样评论丢番图这本书：“在研究了丢番图 100 个解答以后，对于第 101 个问题依然无所措手。”所论或许失之过苛，但不论如何，丢番图一类的问题集，与中国古时以追求普遍适用于解决一大类问题的一般方法为主的问题集相比，是不能相提并论的。

实数系统的建立

实际问题的解答通常需要通过数量来表达。因此，达到这一目标的先决条件是要有一个良好的数系统以及简易的运算工具与运算法则。首先是一个可以表达任意大整数的方法。中国

远古时就为此而创立了完善的10进位值制。世界古时各个民族，都有不同形式不同程度的进位制记数法，如巴比伦的60进位制、埃及与希腊的10进制以及中美与南美马雅民族的20进位制等。但是，它们的进位制有时是不完全的，更谈不上位值制。至于印度，至少在6世纪以前，其以位值制表达的记数方法，即使是个别数字也还没有发现过。当时的记数方法是杂乱无章荒谬绝伦的，不妨参阅印人Datta-Singer的《印度数学史》以及范德瓦尔登(Van der Waerden)《科学的觉醒》等著作。没有一个象样的进位制或即使有但没有完善的位值制，则哪怕是简单的算术运算也是难以完成的。要在这样的基础上建立数学大厦，就更是匪夷所思了。

在我国，由于古黄河流域气候温暖，遍地多竹，人们可就地取材，采取以竹为筹，置筹于盘来进行各种运算。在盘上，不同位置的同型算筹不仅代表某一绝对数值，而且还代表不同的位置数值，由此从10进制进化为具有位值的10进位位值制记数法是颇为自然的。在这种记数法中，自然出现以空格表零的法则，这与后世书写方式中印度之以“◦”或“·”、我国之以“口”和马雅民族之以眼形符号表示零者并无实质上之不同。早在《九章算术》中已经有开平、立方根的算法，其中位值制以及空位作零的作用极为明显，更不用说其他算术运算了。正是由于这种位值制的发明，才使古时中国的数学有可能蓬勃发展，为至少从秦汉至宋元一千数百年间数学的繁荣奠定了基础。

位值制的数字表示方法极其简单，因而也掩盖了它的伟大业绩。它的重要作用与重要意义，非但为一般人们所不了解，甚至众多数学专家对它的重要性也熟视无睹。而法国的数学家拉普拉斯则独具慧眼，提出算术应在一切有用的发明中列于首位。中华民族是这一发明当之无愧独一无二的发明者。这一发明对人类文化贡献之巨，纵然不能与火的发明相比，至少是可与文化

史上我国四大发明相媲美的。中华民族应以出现这一发明而引以自豪。

在整数位值制表达的基础上，我国又逐步扩大数的系统。《九章算术》在方田第一章与少广第四章中详细叙述了分数各种运算与大小比较的法则。在方程第八章中则引进了负数。同在少广章中，又叙述了非位值制无从进行的开平、立方法的详细步骤。在开方不尽时，则附加一分数以为其余，即所谓命分。刘徽在《九章注》中，对开方不尽以及求圆周率的问题提出以 10 进小数逼近的想法。遗憾的是，在刘徽以后的数百年间，这一重要主张未为学者们所认识。但至宋元之际，这种正负 10 进小数已被普遍使用。而在欧洲，则直至 1584 年才由斯蒂文（S.Stevin）（1548—1620）引入 10 进位小数，又迟至 18 世纪才获得通行，并被认为是进位值制之后的又一重大创造。至于西方数学史家历来把这些都归之于印度的发明，则自然是张冠李戴了，这是一种历史性的错误，在此不能不辩。

整数位值制的建立，负数以及分数与小数的引入，使整个实数系统得以完满地完成。而有无实数无理数之类的概念与名称，则是另一问题。对于实数的现代认识，本来还是 19 世纪中叶才有的事。完善而使用方便的数系统，创造了十分优越的条件，使我国的传统数学一千数百年间大放异彩。

数论的辉煌成就

中国从来没有素数与分解因子的概念，但这并不妨碍我国以自己独有的方式发展数论。

中国虽没有素数与分解因子的概念，但有最大公因子的概念及其求法。只是当时称两数的最大公因子为“等”，求法是“以少减多，更相减损，求其等也”。这一方法见于《九章算术》方田章的约分术并不是偶然的。现代为了求分数之和而先通分时，

往往先将诸分母分解成素因子的乘积，但是分解因子在整数较大时并非易事，相反求“等”即求两数的最大公因子则要容易得多。《九章算术》少广术就避开了分解因子这一步，巧妙地只利用求等得出了分数之和，这比现行方法要简便许多。

勾股形(即直角三角形)的三边何时成整数比，自然是数论中颇具代表性的问题。《九章算术》勾股第九章中出现这样可能的分数比有 8 组之多。由求等法容易化分数比成整数比，即刘徽《九章注》中所谓“如是或有分，当通而约之乃定”。出现这样多的整数比决非偶然，事实上勾股章已提出了构造这种比的一般方法与一般公式，刘徽《九章注》中还给出了公式的几何证明。至于中国以外，则上述问题最早是在 12 世纪也以分数比的形式出现于印度婆什迦罗(Bhaskana)(1114～1185 年)的著作中。

我国在数论上的又一杰作是大衍求一术，即现代通称的中国剩余定理，用现代术语相当于一组同余式的解法。大衍求一术初见于宋时秦九韶 1247 年的《数书九章》一书，但其来源则是天文历法中的上元积年计算问题，且从汉以来历代的历法中均有一条颇为清晰的发展路线。至秦书时则不仅有详细的算法，还把应用推广到历法计算以外的河堤修筑。财物计算、租税分配、谷物出售、军队计点、土木建筑等问题上，甚至还有一个盗窃追赃的案例。现代解决这一同余式问题通常限于各同余式诸模数两两互素的情形。方法是先将诸模数分解成素因子的乘积，并求出其欧拉数，最后得出所求的解。因为分解因子比较繁难，所以对于秦书诸题中出现的天文数字般的大数，这种方法由于计算量的庞大，即使用现代的计算机要完成其计算也是不容易的。对于根本没有素数概念的中国传统数学，秦九韶却利用了求等的方法，首先将模数的一般情形化成模数两两无等即两两互素的情形，再用与求等类似的方法巧妙地得出其解法，这在世界数学史上也是很了不起的一举。当时计算用等，即使如此，对

于有着天文数字般大数的各个问题，也都轻易地获得了答案。秦书中并记载了所有计算过程的详章。

（二）

代数学的蓬勃发展

代数学是中国传统数学中又一个值得骄傲与自豪的领域。与现代理解的代数学有所不同，它完全可以等同于方程求解。而且，即使是方程，着眼点也是求出解答，至于解答的存在唯一、解答的性质、解答与系数间的关系以及解答可否用系数的开方式来表示一类问题，都非关注所在。

早在《九章算术》中，第八方程章全章就是线性联立方程组的解法。依刘徽注："并列为行，故谓之方程。"又说："令每行为率，二物者再程，三物者三程，皆如物数程之。"当时以筹作算，依题意将数据按物数排成各行，然后进行运算。《九章算术》中还载有某些具体实例的详细计算过程，完全类似于现在的所谓矩阵运算与消去法。列方程时需要相当于移项一类的手续，因此引进了负数以及相当于正负数加减运算法则的正负术。刘徽在注中又引入了新术与又术，对同一问题还指出了不同方法的不同计算步数以相互比较算法的复杂程度。中学教科书解线性方程组的消元法，最早即见之于《九章算术》。

在现代，这种解法通称为高斯消元法。它的最早来源是高斯在 1826 年所著的关于天文观测误差估计的一篇文章，文中提到的线性方程组具有特殊形式。此外，Cramer 在 1750 年也曾发现过后来以他命名的公式。有趣的是 Cramer 发现在前而高斯在后。而且，在计算机发展的早期，一位计算机科学家曾撰文举例指出：在计算机上求解一个 26 个未知数的线性方程组，如果用 Cramer 方法需时将近 10^{17} 年，而用消去法则 3 秒钟就够

了。以 Cramer 法则与中国传统数学中的消去法相比较，一尚优美，一尚实效，两者旨趣各异，代表了两种不同的风格。这样的例子并不少见。

从勾股术求弦需要开平方，这相当于解一个最简单形式的二次方程。在中国古籍中历来称解方程为开方，可能即导源于此。《九章算术》中不仅有开平、立方的详细算法，并有一题涉及较一般形式的二次方程，称为开带从平方。约在三国时期赵爽的现存残文中即有解一般二次方程的痕迹。南北朝祖冲之，按《隋书》有开差幂、开差立之设，称为“算氏之最”，有可能是开带从立方即解某种形式三次方程的方法，可惜已完全失传。唐初王孝通的《缉古算经》则是关于三次方程的专门著作。我国传统数学以解决问题为主要目标，对方程的要求是求出数值解答，而不在形式上的一般公式。而后者据近代的数学研究本来就是无法实现的。王孝通的著作在说明如何对诸问题列出三次方程后，求具体答案只用“开立方除之”一语带过。由此可以想见，三次方程的数值求解问题已为当时仕人所熟知，故已不必多加辞费。此外，在历法与河防的典籍中，也出现过需要求解三次或四次方程的问题。至宋代，出现贾宪“增乘开方”、刘益“正负开方”等术，到秦九韶已完成了一般方程的数值解法。笔者在袖珍计算器上依据这一方法解高至五次的方程，其效率是极为显著的。

宋元时期天元一概念的引入以及天元术的创立，是我国传统数学的又一重大成就。天元一即是现在的未知数。这一概念的引入提供了依题意建立方程的简便而系统的方法，较之旧法“省功数倍”。由此逐步发展至元朱世杰的四元术，已可解多至四个未知数的高次联立方程组。由于当时用筹算，方程各不同类型项的系数须布置在算盘的特定位置来进行运算，因而未知数的个数只能限于四个。如果改用纸笔运算，则四个未知数的

限制完全可以打破。至于原来四元术的原理与方法，则仍可以适用于解任意多个未知数的高次联立方程组。

在近代，解高次联立方程组的方法直到近一二十年才有重大进展。据笔者所知，在去年曾出现过一些文章，对当前某一最有成效的方法加以改进，并在计算机上进行了试验。其中至少有两个经过不同改进看来很简单的方程组，竟然在计算器上有一个未能解出，另一个也要动用计算机很大容量和花费很长时间才能勉强算出。笔者依据本人创立的方法，对这两个方程组在一些小得多的机器上即可容易地求解。这里所谓本人创立的方法，实质上无非是朱世杰四元术的现代化推广形式。上述两个方程组论次数不过二三次，论未知数或方程个数也只有三个(附带若干个参数，其中一个方程组只有一个参数)，鉴于朱世杰著作中就出现过有三个未知数高至三次的方程组，可以想见中国古代算家着眼解题，重视实效，其思想之深邃与效率之高是为一般人所难以想像的。

别具一格的几何学

几何学乃是中国传统数学中最不为人所理解的一个领域。其根本原因是人们对欧几里得几何已经先入为主，而我国古时的几何却与欧几里得几何大相径庭，有着一个迥然不同的体系。

与欧几里得几何不同，我国传统几何从来不考虑平行线的问题。与之相反，以垂直性为其特征的勾股形，在我国古代几何学的研究中却始终占据着中心位置。数学概念来自现实世界。过直线外一点可以作几条平行线的问题，至少在当时的现实经验看来是近于“无病呻吟”毫无意义的。对于重视解决实际问题的中国古代算家来说，完全排除这一类考虑是理 所当然的。

与欧几里得几何不同，我国古代几何极少考虑角度，而把距离与长度放在首要地位。如果考虑到现代黎曼几何的从无限小距离 ds^2 出发，可以导出角度的概念与性质，则中国古时把距离的位置置于角度之上也就不足为奇了。

对欧几里得几何来说最有代表性的是它首次建立了在公理系统基础上的演绎体系，成为现代数学的一种典范，这在多方面具有重大意义。但也不能不注意到，这种体系与方法即使在纯粹数学的研究中，其作用也是颇有局限性的。就是对几何学来说，公理系统的建立也基本上局限于初等几何。对于现代研究中极为活跃的代数几何与微分几何，并没有什么公理系统。拓扑学虽然出现过某些成功的公理系统，但拓扑学中的重大成果，并非单纯地从这些公理系统经过逻辑推导而得。现代数学家们已经严格证明，要单纯地从某一公理系统出发去 证明系统中的定理，不仅实际上难以做到，在理论上也无此可能。

中国的传统几何历来遵循着与欧几里得几何完全不同的发展道路，有着自己的问题与方法，以及自己的理论体系。

首先，我国古代几何重视的是实际问题的解决，因而测量、面积、体积以及圆周率等都是几何学研究的中心问题。对于圆周率知者甚多，可以置之不论。对于测量，则由于我国古时重距离而不重角度，因而所用仪器与方法与西方大不相同。至迟在汉初，我国即已出现用距离两测以测角的方法求得太阳高度这样的困难问题。至三国刘徽，更发展成一般的重差方法，所著《海岛算经》包括了多种复杂测量的漂亮公式。即使在现代，要设计某种测量方案以获得答案，恐怕也不是那么轻而易举的。

我国古代几何成就中最值得骄傲的是体积理论。欧几里得体系以许多定义与公理为出发点，然后据此进行逻辑推理求体积。与之不同，我国古代几何着眼于总结经验，综合事实，提炼

出聊聊几条普遍而平凡的一般原理，然后用逻辑推理推导出多种多样求体积的结果来。这种方法正与整个经典力学可以建立在三条牛顿定律上者相类似。在这种思想指导下，在三国时期刘徽与赵爽的著作中即已出现了出入相补原理。据此即可建立测量理论与多角形的面积理论。开平、立方的算法以及勾股形三边整数比公式的证明也都奠基于这一原理。《海岛算经》诸公式形式奇特，其证明早已失传，但依据出入相补原理可以轻而易举地推导出这些公式。如果用那种添平行线甚至用三角术等方法来证明，则不仅违背历史，尽失古意，还要费不少心机才能迂迴曲折地求得那些公式。

多面体的体积比之多角形的面积是一个艰深的问题。依据现代的数学研究，当时要建立体积理论，不可避免地要引入相当于今日极限、连续这类的概念，因之与面积理论有本质上的区别。我国古代算家敏锐地认识到这一点，并由刘徽予以彻底解决。刘徽依据任意多面体由某些基本形体阳马、鳖臑等拼合而成，并从长方体的解体剖析，通过极限过程建立了“阳马居二、鳖臑居一，不易之率也”这一二比一原理。这一原理与出入相补原理相结合，即可奠定多面体体积理论的基础，并正如《九章算术》所示，可容易地获得各种多面体的体积公式。整个理论简单优美，与古希腊处理同类问题用穷竭法之拖沓繁琐者相比，优劣判然。这是我国古代几何学的一项杰作。

曲体特别是球体的体积，显然不能仅仅用上面两个原理来解决。刘徽指出了某一特异曲体所谓“牟合方盖”在这一问题中的作用以及解决整个问题的途径。后终于在南北朝时由祖冲之之子祖暅提炼出“缘幂势既同，则积不容异”这一原理以作为求曲体体积的基础，并且彻底解决了球体体积问题。我们称这一原理为刘祖原理。至17世纪时，它又为意大利 Cavalieri 所重新发现而被称为 Cavalieri 原理，成为微积分得以建立的基石之一。

结 束 语

数学是一门研究现实世界中的空间形式与数量关系，即研究数与形的科学。几何学致力于形的研究，因而也可称为形学。欧几里得几何脱离形来自现实世界的实质，而着眼于抽象出来的概念与性质及其相互间的逻辑关系。与之相反，我国古代几何的对象都直接来自现实世界，问题也大都来自实际需要。欧几里得几何形数脱节，实际上排斥了数量关系。与之相反，在我国几何学的整个发展过程中，始终形数合一，互助互益。开平、立方与解二次方程，都以出入相补原理为其几何背景，即其一例。

宋元之际，引进了天元一的概念。各种几何量都用天元一的格式表达。这种表达式相当于现代的多项式。几何问题变为求解方程组的问题。四元术更发展了导源于《九章算术》方程术的消元法，使多至四个方程组可逐步化为只有一个未知数的方程以逐一求解。这种几何的代数化为解析几何的出现迈出了重要的、也是决定性的一步。

解析几何的创立者 Descartes，有一篇直到死后才刊印的未完成著作《指导思维方向的法则》(《Rules for the direction of the mind》)。在这一著作中，Descartes 拟议了一个解决问题的普遍方法，期望它能适用于一切类型的问题。其方法大致如下：第一步，把任一问题化为一数学问题；第二步，把任一数学问题化为一代数问题；第三步，把任一代数问题化为一解单独一个方程的问题。Descartes 还提出应按照未知数的个数建立同样多的方程，而每一方程都是某一多项式等于零的形式。显然，Descartes 的计划过于宏伟也过于粗疏，事实上无法做到。可能正因为如此，Descartes 没有完成他的《指导思维方向的法则》这一著作，但却从这一宏伟计划中抽出一部分来写

成了他的不朽之作《方法论》，并在附录中阐述了现在被称作解析几何的概念与方法。顺便指出，这一附录中还没有引进坐标，线段也不分正负，因而实质上只是一种有系统的几何代数化方法，至于 Descartes 这些著作对整个科学特别是数学的重要影响，则已是尽人皆知的事了。

回顾我国从秦汉到宋元间数学发展的历程，可谓我国传统数学所走过的道路正好与 Descartes 的计划若合一契；反过来，Descartes 的计划，也无异于为中国传统数学作了一个很好的总结。

中国传统数学的成就决不止于上述内容，但由于篇幅所限，更由于笔者知识所限，对其他成就将不再论述。

我们崇拜中国传统数学，决非是泥古迷古，为古而古。复古是没有出路的。我们的目的不仅是要显示中国古算的真实面貌，也不仅是为了破除对西算的盲从，端正对中算的认识，我们主要的也是真正的目的，是在于古为今用。

数学在中国古时历来称为算术。这正好反映了中国古算构造性、计算性与机械化的特色。数学不仅为了应用，即使为了在纯数学范畴内获得具体结果，也一不能无算，二必须有术。对于概念与概念之间逻辑关系的认识，不仅在过去而且在将来都起着重要作用。但也不能不有所警惕，追求过甚不无有陷入推理游戏甚至数字游戏的危险。算术两字，决非对数学的卑词。试对 17 世纪数学上的两大成就解析几何与微积分略加分析：解析几何为使欧几里得式的几何证明化为中国传统数学的计算方式开辟了道路，无非是把几何化为算术的一种方法。微积分按西文名词不妨直译为无穷小算法。在牛顿与莱布尼茨之前作为微积分的先驱者瓦尔利斯（Wallis）的一部著作，其名称直截了当就叫做《无穷的算术》。由于计算机的出现，算术化的倾向在近代数学中的作用也日益显著，越来越为人们所认

识。不少著名的数学家，近年来纷纷转向于计算机代数、计算性几何一类新兴学科。中国古代算术的思想与方法，正好与近代计算机的使用融合无间，也必将因此而重返青春，以另一种崭新面貌在未来的数学发展中扮演重要角色。究竟它将占据何种地位，笔者不拟妄作预测，但愿以二语作结：事在人为，还看今朝。

《现代数学新进展》序*

1985年10月4日至13日，中国科学院系统科学研究所在部分同志倡导下，举办了以刘徽命名的数学讨论班，邀请十几位数学专家对现代数学的某些领域及其成就作了概括性报告。本书就是这些报告的一个综合。

这一讨论班的出现受益于驰名国际的Bourbaki讨论班的启发。

本世纪30年代，法国的一些青年数学家创立了Bourbaki学派。学派的主要创建人如A. Weil, H. Cartan, C. Chevalley, T. Dieudonné, J. Delsarte, Ch. Ehresmann等，都是当时法国培养数学家的中心和基地——巴黎高等师范学校的学生。这些青年后来都在学术上取得了杰出的成就，从而成为国际数学界的巨擘。但是，他们的活动并不局限于个人的学术研究。他们以Bourbaki为集体，举办了若干对全世界数学发展有重大与深远影响的活动。其一是《数学原理》全书的编写，其二是Bourbaki讨论班的设立。

数学经过几千年的发展，已成为分支庞杂的宏伟体系。学者终其一生，即使是一流大师，也往往只能在少数领域有所建树，很难了解数学全貌。Bourbaki集体提出了用结构这一概念来贯串整个数学，并着手编写《数学原理》，从无结构的集合论与具有最基本结构的实数论开始，依次进入结构不同逐步丰

---

* 本文摘自《现代数学新进展》(吴文俊主编)，安徽科技出版社，1988年。

盈的各个领域。该书各分册由 Bourbaki 中成员分头执笔，但是必须经过集体讨论，集体修改，并以 Bourbaki 署名。每本书的编写，往往数易其稿。至今已出版数十分册，历时四五十年，但尚不能预料完成的时日。编写工作也已由 Bourbaki 中老的一代交卸给新的一代。这部鸿篇巨制不仅对数学的发展有巨大的影响，而且给法国数学界带来了极高的声誉。

博与精难以得兼。数学家为自己所从事的课题研究，已耗费了大部分精力和时间，对与研究课题无关的领域，往往无力涉猎。Bourbaki 讨论班之设立恰好弥补了这一缺憾。这一讨论班实质上是一种数学动态讨论班，报告的内容并非个人的研究成果，而是介绍国际上当前某些重大发现。该讨论班一年举办三次报告会，每次约三天；在报告会的中间提出下一次值得介绍的课题，并由与会者自告奋勇地去准备，有时也邀请外人作报告。报告人在报告时往往融合自己的思想和创见。由于其内容的精辟，影响已远远越过了法国国界。历届讨论班都编印报告论文集刊行，成为数学上创新的重要源泉，为全世界各个不同领域中的数学家共同的重要参考文献。笔者 50 年代初参加这一讨论班时，全班不过 20 来人，在巴黎高等师范学校的一个小教室中举办。1982 年重游巴黎时，地点已移到巴黎 Poincaré 研究所的一个阶梯教室。笔者到达时虽不算晚，但不仅座无虚席，而且阶梯过道上也坐满了人，后来者只好立于门外。许多人是专程从远道赶来参加的。

Bourbaki 学派对青年一代的培养极为重视。姑且不说老一辈中如 Weil，Cartan 曾获得可与 Nobel 奖相埒的 Wolf 奖，在他们培育之下成长的如 L. Schwartz，J. P. Serre，R. Thom，A. Grothendieck，P. Déligne 等，都先后获得了在数学界享有最高荣誉但只授予青年数学家的 Fields 奖。

50 年代以来，Bourbaki 的影响已波及整个数学界。青年

数学家纷纷将 Bourbaki 奉为圭臬，以《数学原理》为学习基础，钻研 Bourbaki 学派的著作，追随他们提出的研究方向，接受他们的结构思想，推行他们倡导的公理化体系。

这些虽然都是 Bourbaki 学派的伟大业绩，但还仅仅是其外部表现，而不足以说明其精神实质。

笔者在国外曾遇到一位第三世界的数学家，他说了这样一句话："Bourbaki 是法国民族精神的产物。"

此语可谓一针见血。这位数学家口中的 Bourbaki，才是真正的 Bourbaki!

文艺复兴与资产阶级大革命以后，法国已成为欧洲科学文化的中心，数学界尤其人材辈出。拿破仑在执政时又创办了独具一格的多工艺学校，由数学家 G. Monge 主持其事。从此多工艺学校就成了法国培养数学家的中心与基地，直到 J. Douboux 活跃于数学界时才让位于巴黎高等师范学校。从 17 世纪 R. Descartes, P. Fermat，牛顿，莱布尼茨等创立解析几何与微积分起，法国在长达两百多年的时期一直执欧洲数学发展即世界数学发展的牛耳。但是到 19 世纪中期，德国数学崛起，逐渐后来居上，数学中心有从巴黎转至德国的 Göttingen 与柏林之势。进入本世纪后，与德国相比，法国数学研究的范围日益显得有些偏狭，更少新的思想。函数论向来是法国数学的一张王牌，但到本世纪 30 年代，J. Hadamard, E. Borel, E. Picard, P. Montel, G. Julia, G. Valiron 等的光辉成就已成强弩之末，难以为继，他们的方向也不再象过去那样在整个数学王国中占据核心地位，法国数学已濒临丧失过去二百多年来国际领先地位的境地，而且与周围各国的差距颇有扩大之势。在这样的形势下，法国一些年轻而有才华的有心人创立了 Bourbaki 学派，经过数十年的惨淡经营，终于使法国数学重新占据世界舞台的中心。在构造概念下对全部数学的

统一处理，《数学原理》全书的编写，Bourbaki 讨论班的创立，对青年一代的培养，凡此种种，都无非是在以复兴法国数学为历史使命这一指导思想下产生的数学思想与具体措施。当年都是二十多岁的年轻人，如今都已耄耋老矣，有的已经故世。近年来，Bourbaki 的影响已见衰退，对他们的思想与体系也颇有争议，并不时受到非议，其成功确也有一定的范围和局限性。但是他们为重振法兰西精神所作的努力，不仅对法国人民是可贵的，也可供其他各国人民借鉴与学习。我们要向 Bourbaki 学派学习的，不在于他们在各个领域取得的各项特殊的成就，也不在于他们时有争议的思想体系。这些都在可学可不学、可从可不从之间。真正值得我们学习的乃是他们这种可贵的精神。

中世纪的欧洲，一直为封建领主所割据而分成无数小块领地。至 10 世纪末，法兰西虽然名义上成为统一的国家，实际上仍然四分五裂，直到英法百年战争之后，于 15 世纪才结束了封建割据，实现国家统一至今不过四百多年。与之相较，早在公元前 211 年，秦始皇统一六合，中国就成为中央集权的封建大帝国。自秦汉迄宋元时代，我国在数学上绵延不绝的光辉成就，为现代数学打下了基础，也为数学的未来发展做出了楷模。只是自明代以来的几百年中，我国的传统数学骤然衰微，几乎退出了历史舞台。振兴中华，对数学工作者来说，不仅是振兴的问题，而且还有一个复兴的问题。Bourbaki 学派复兴法国数学所作的努力，为我们提供了一个良好的榜样。这是我们创设刘徽数学讨论班的缘起。

发起的同志认为讨论班有必要冠以某一著名数学家的名字，原拟为祖冲之讨论班，最后确定为刘徽讨论班。

祖冲之，5 世纪南北朝人，最为国内外知名的工作是关于圆周率 π 的计算。《隋书·律历志》载：“宋末，南徐州从事史

祖冲之更开密法。……密率：圆径一百一十三，圆周三百五十五。约率：圆径七，周二十二。”用现代的形式描述即

$$\frac{22}{7}>\pi>\frac{355}{113}.$$

此外，祖冲之最主要的著作是《缀术》，号称难读，并已失传，内容已不得而知。他的儿子祖暅，为解决长期悬而未决的球体积问题提出“幂势既同，则积不容异”的原理，也就是迟至17世纪又被重新发现的 Cavalieri 原理。众所周知，这一原理是微积分得以产生的主要推动力之一。祖冲之不仅是一位伟大的数学家和天文学家，而且也是一位伟大的工程师。他曾制造过指南车、欹器、千里船、水碓磨等机械，经试验都很有效，可以说是我国古代一位 Leonardo da Vinci 式的伟人。西方撰写的数学史提到我国古代的数学家时，常以祖冲之为代表。鉴于祖冲之在科学技术上的成就，受到国内外如此尊崇，应该说是很自然的，而且是无可非议的。

但是，从数学的角度来说，祖冲之不能视为我国古代数学史上的代表人物。真正的代表人物应该是刘徽，而不是祖冲之。

刘徽，三国魏晋时人，生卒年月不详。我国古代数学的经典代表著作是成书于公元前后100年间的《九章算术》（以下简称《九章》）。刘徽的主要著作一是为《九章》作注（以下简称《九章注》），另一是原拟作《九章》新补一章但后来单独成书的《海岛算经》。对于从事中国古代数学史研究的同志来说，把刘徽看成我国古算的代表人物应该是无庸置疑的。但对于一般人来说，刘徽可能是不见经传的人物，与祖冲之之家喻户晓不能相提并论。

笔者认为，一般人，包括很大一部分数学工作者，只知有祖冲之而不知有刘徽，其中颇有缘故。不论中外，圆周率的计

算历来都是众所瞩目的问题。在圆周率计算上有突出成就的祖冲之，也就容易得到对我国古代数学颇为隔膜的西方数学史家的赞扬。追本溯源，很可能是由对我国古算仅仅一知半解的西方传教士先入为主的介绍所致。不论有意无意，真正代表我国古代数学的《九章》，西方传教士几乎从未提到。近代，《九章》虽然也被翻译成几种外文，但其真正精髓《九章注》，至今仅有一种日文译本。我国古代数学，至明季已几成绝学。现代的数学，自明末西方传教士进入我国开始，完全从西方引进。一般人对我国传统数学的认识，也就往往以西方数学史家的著作为依据，西方传教士的一孔之见，不知不觉地深入人心而成为普遍的看法。

祖冲之父子的主要著作《缀术》早已失传，其内容不得而知。因而对祖冲之在数学上的成就很难作出全面确切的评价。就圆周率计算而论，虽然祖冲之的疏密二率是一项杰作，但二率的得来各家说法不一，颇难臆测；估算较疏但早于祖冲之者国外有阿基米德，国内据《隋书》所载就有刘歆、张衡、刘徽、王蕃、皮延宗等，晚于祖冲之但估算更精密者则有15世纪阿拉伯的阿尔卡西等。因而祖冲之在数学史上的地位仅凭这一工作并不能评价过高。相反，计算圆周率的理论根据需要某种极限的概念。通过圆内接多边形周界极限来计算圆周率的方法，刘徽在《九章注》中已作了详细的解释。在没有发现其他文献可以印证的情况下，刘徽无疑应视为圆周率计算理论与方法的真正奠基人与缔造者。同样，祖暅原理即后世的Cavalieri原理虽然出自祖暅，球体积的计算也完成于祖暅，但在刘徽的《九章注》中，早已有了这一原理的痕迹，刘徽并且已应用于某些简单曲体体积的计算，只是没有形成文字而已。把这一原理改称为刘祖原理，亦无不当。至于祖暅的球体积计算，依赖于某一古怪的立体所谓牟合方盖体积的计算。但牟合方盖以及

由此通往球体积计算的道路，刘徽在《九章注》中都已指明。自刘徽至祖暅的数百年间，不妨可以认为正是由于数学家们孜孜矻矻地遵循了刘徽指明的道路，才终于完成于祖暅原理的。

《九章》特别是刘徽的《九章注》，是我国传统数学的伟大宝库，是直至宋元时期我国在数学上许多重要发明创造的源泉。刘徽对数学的贡献，足可与古希腊的贡献相提并论，对现代数学的影响，也决不在古希腊的影响之下。以上所举的圆周率与球体积计算，只是刘徽众多贡献中的两例而已。我们不拟对刘徽的贡献作较详细的介绍。读者尽可求之于刘徽的原著或有关中国数学史家的专门著作。

为《九章》作注者并非刘徽一人，在刘徽前后都不乏其人。但除了唐代李淳风作的补注以外，只有刘徽的《九章注》流传至今。我们不妨认为《九章注》是刘徽以及其前与同期我国数学家聪明智慧的结晶，而以刘徽为这些古贤哲的代表，正象 Bourbaki 是一个集体的代表名称那样。在这种意义下，刘徽无可争议地是我国传统数学中唯一的代表人物。

对于我国的中国数学史专家来说，刘徽之为我国传统数学的代表人物本来是一种常识。但我国传统数学濒于失传并让位于西方近代数学已有几个世纪之久。因而我国的数学家容易以对我国传统数学几乎一无所知的西方数学史家的舆论为舆论，而忽略了熟悉我国传统数学的我们自己数学史家的真实意见。这可能是历来盛称祖冲之而刘徽之名不彰的重要原因之一。

这一轻重倒置与我们创设这一讨论班的主旨根本不相容。因之，这一讨论班不能冠以祖冲之之名而应以刘徽来命名。这就是刘徽讨论班名称的由来。

这次讨论班由 15 位专家分别对数学中若干重要领域、重要理论与课题作了概括性的介绍。他们的报告除了一篇因未交稿以外都收在这本文集中。所介绍的领域有数理统计、线性规

划、模型论、复几何、Kac－Moody代数、非标准分析、模糊数学等。报告人都是有关方面的专家并有他们自己的贡献。这些领域有些历史悠久，其重要性久已为人们认识，有些虽然出现较晚，但影响颇为深远。例如50年代开始提出的模型论，是60年代出现的划时代的非标准分析的基础。又如线性规划，虽然出现于60年代初期，近年来由于出现了苏联Khachiyan和印度Karmarkar提出的新方法引起了重大反响，《参考消息》上也有过多次报道。再如复几何，由于陈省身、丘成桐等的重要贡献，成为纯数学中当前极为活跃的一项中心课题，报告人钟家庆同志由于这方面的工作而获得第一届陈省身奖。某些领域形成未久，例如Kac－Moody代数，因1968年Kac的工作而得名，由于在物理学及其他方面的应用而得到迅速发展。另外，某些领域如模糊数学，历来虽有争议，但其应用前景颇为乐观，有待于在这方面已有显著成就的我国数学家继续努力。又如非标准分析，于60年代初为A. Robinson创立后，数学家们对之持不同态度，但近年来由于法国Strasbourg大学G. Reeb领导下数学家做了出色的工作，Bourbaki讨论班曾作专门介绍，局面势将改观。我国数学家已率先将非标准分析成功地应用于解决广义函数乘法问题，这是广义函数问世以来即被公认的难题。

近年来，众多长期成为悬案的难题获得了重大进展，以至彻底解决。例如，关于Bieberbach猜想的论文历来数以万计，却为de Branges一举获得完全的证明。尤其难能可贵的是组合数学中存在了百多年的所谓Steiner三元系问题，直至1980年还认为解决无期，却在1981～1983年间为我国内蒙古一位中学教师陆家羲几乎完全解决。在他不幸早逝时，只留下了六个孤立的例外。陆家羲同志由于这一杰出成就获得第三届全国自然科学奖一等奖。此外如气体动力学中的Riemann问题，已

有百年以上的历史，向为 Courant 研究所的传统研究课题，我国丁夏畦、罗佩珠、陈贵强等同志对此取得了重要的突破性的进展。陈贵强同志因此被 Courant 研究所邀请访问已有两年之久。中国科学院已初步评定授予丁夏畦同志科技进步奖特等奖。又如不动点理论向来是拓扑学与应用联系的一个重要纽带。其中关于不动点几何个数的 Nielsen 理论出现于本世纪 20 年代，我国数学家江泽涵、姜伯驹、石根华等同志作了重要的发展。Nielsen 的一个重要猜测直到数年前才为姜伯驹所解决。

有些理论与方法是我国数学家所独创的。例如廖山涛同志关于微分动力体系的理论，获得了 1986 年度第三世界科学院奖和我国第三届自然科学奖一等奖。又如洪加威同志的几何定理例证法，思想新颖，不仅在数学上而且在哲学思想上发人深思，已在国际上引起轰动。

我国传统数学有其自身的发展途径与独到的思想体系，而以机械化为其特色；方程求解尤其是贯穿两千多年发展中的一条主线。这与遵循古希腊传统的西方数学的公理化演绎体系大相径庭，旨趣迥异。在历史长河中，数学机械化算法体系与数学公理化演绎体系曾多次反复互为消长，交替成为数学发展中的主流。肇始于我国的这种机械化体系，在经过百年来的消沉后，由于近代计算机的出现，已越来越为数学家所认识和重视，势将重新登上历史舞台。笔者关于几何学的机械化方法，本质上即直接导源于我国传统数学的思维方式，目前在几何学定理的证明上，已获得极为可观的成功。这一方法同样可用于方程求解，这也是对我国传统的直接继承。由于理论问题与实际中来源不一的形形色色问题，最后往往归结为求解某种类型的方程，因而其应用前景极为宽广。我们的方法为复兴我国的传统数学，提供了一个有效的切实可行的途径。笔者在这方面的第一篇著作《初等几何判定问题与机械化证明》，刊载于

1978年《中国科学》。笔者在该文之末加了一个附注，现将这一附注照录如下：

> “我们关于初等几何定理机械化证明所用的算法，主要牵涉到一些多项式的运用技术，例如算术运算与简单消元法之类。应该指出，这些都是12至14世纪宋元时期中国数学家的创造，在那时已有相当高度的发展。详细介绍可参阅钱宝琮的著作（中国数学史，科学出版社，1964）。事实上，几何问题的代数化与用代数方法系统求解，乃是当时中国数学家的主要成就之一，其时间远在17世纪出现解析几何之前。”

讨论班的创立曾得到北京中关村科技发展公司的赞助，在此顺致谢意。国外近年来有了重大发展而国内又有这方面专家可作深刻分析介绍者为数众多，国内也还有不少发明创造。由于条件限制，讨论班的报告人只能局限于北京及其近邻地区。还有些同志由于出国访问或其他原因，虽曾邀请作相应的报告但未能如愿。不过，仅从收集在本文集的报告内容来看，已足见我国在数学上已具有一定的实力和强大的潜力。复兴而不仅是振兴中国数学，使自秦汉迄宋元傲居世界舞台中央的中国数学重展昔日雄风于今日，应该是完全可能的。

《陈省身文选、传记、通俗演讲及其它》序*

科学出版社决定出版《陈省身文选》，内容包括陈省身教授的许多通俗演讲、综合报告、著作与人物评介，以及对自己的传记文字等。出版社要我写一篇序，并把《文选》几乎全部文章的复印件交给我，以作参考。这使我感到无上荣幸，又感到难以胜任。但在将这些复印件翻阅之后，使我回想起1946至1947年在中央研究院数学研究所期间，在陈师指导下学习拓扑学的种种经历，故作此随笔，以志不忘。

我在国外访问期间，曾与国际友人谈起各人的学术经历。我说起我与陈师本不相识，只是在中央研究院数学研究所耽了一年，从陈师学习代数拓扑，从此走上了拓扑的研究道路。闻者大为惊异，拓扑号称难学，一年就在拓扑上做出研究成果，认为不可思议，因而见人就说此事。其实这并不可怪，这正好说明陈师善于提携后进，指导有方所致，如此而已。

经过是这样的，陈师是清华大学也是西南联大的教授，而我毕业于上海交通大学数学系。时值抗战，我常年蛰居上海，对外界数学情况颇为茫然，对陈师也一无所闻。1945年抗战结束，我有暇得以复习旧日所学的数学。与陈师相识，全靠亲友帮助介绍。其时陈师自国外回上海主持中研院数学所，经朋友介绍往见陈师。亲戚并为我打气，说陈先生是学者，只考虑

---

* 本文摘自《陈省身文选》，科学出版社，1989年。

学术，不考虑其它，不妨放胆直言。在一次与陈师晤谈中，我直率提出希望去数学所。陈师不置可否，但送我出门外时，却说：你的事我放在心上。

过了没有多久，陈师通知我去所工作，从此我便走上了数学研究的道路。

当时的数学所规模很小，只占据一座楼的第二层。最大的一间供会议与报告之用，次大的是图书室。我被安排在图书室作为工作地点。陈师独居一室，只记得有一架打字机，陈师经常在上面用一个指头打字。其余大都是大学毕业未久的年轻人，分居各室。我到那里时数学所刚成立，陈师出身北方大学，但对吸收年轻学子毫无门户之见。他们来自武汉大学、浙江大学、上海大同大学，我来自上海交大。来自西南联大者只有陈国才一人。

数学所只办了 3 年。在将近 40 年后，1985 年陈师又在天津办起了南开数学所。两个数学所虽然人物已非，内容有异，但都体现了陈师的宏伟意图，想通过它们来振兴中华数学，使中国在未来成为与国外平等独立，甚或领导世界的数学大国，有步骤有计划地稳步进行，前后是颇为一致的。南开的数学所，正是 40 年前中研院数学所不幸中断的一个继续。

中研院数学所的第一年，我们的学习集中于代数拓扑，陈师为此一周要讲多达十二小时的课，并经常到我们的房间里来讨论拓扑中的各种问题。在这一年中，陈师很少讲到微分几何。我在数学所只耽了一年，以后数学所搬往南京，又新来了不少人，也仍以代数拓扑为研究与学习的中心。但在私下里，陈师曾多次和我谈起，他的主要目标不是拓扑而是大范围或整体性微分几何。

E. Cartan 是近代最伟大的微分几何学家[1)]。陈师是 E. Cartan 的当之无愧的继承人[2)]，也是现代微分几何的奠基人。

E. Cartan 的全部著作中的微分几何部份，几乎全部局限于局部性的微分几何。虽然在晚年注意到李群的整体性质，并提出后来为 L. S. Pontrjagin 所证明关于古典李群 Betti 数的可能公式，以及后来为 de Rham 所证明对微分流形拓扑性质带有根本性的猜想，但本人并非拓扑专家，且垂暮之年也已无力为此。代数拓扑虽创自法国的 H. Poincaré，但直到 30 年代，法国并没有真正的代数拓扑学家，法国第一个这样的拓扑学家，是 E. Cartan 的学生 C. Ehresmann。他为了完成博士论文需要拓扑学，曾在美国普林斯顿耽过一年，就学于 S. Lefschetz 等。虽然如此，在 E. Cartan 的著作中，既指出了拓扑学对于微分几何发展的美好前景，又蕴含了许多对于拓扑学本身极有重要意义的精邃思想。C. Ehresmann 就在 E. Cartan 著作的启发之下，引进了纤维丛与联络的一般概念，成为纤维丛理论与近代联络论的奠基人之一。但更重要的发展则无疑来自陈师。

陈师在四年一次的国际数学家大会上，先后作过三次报告。第一次是在 1950 年，是一小时全会报告[3)]。第三次在 1970 年，也是一小时的全会报告[4)]。在 1970 年的报告中，陈师指出，“除了少数孤立的结果外，大范围微分几何一直等到代数拓扑和李群为它铺平了道路后才得到发展”，而“大范围微分几何是一个年青的领域”。事实上，使大范围微分几何从少数孤立的结果得以蔚然形成当前最活跃的独立分枝之一者，可以说正是陈师本人。纤维丛与联络的概念虽然早已隐含在 E. Cartan 的著作中并由 Ehresmann 与陈师提炼出来，但陈师与 Ehresmann 不同之处是：后者只对概念提出了明确的描述，而前者则不仅如此，还提出了从事这方面定量研究的方法、工具与实例——即示性类，特别是引入以陈师命名的陈类，示性类在联络之下的具体表达式，以及 Gauss－Bonnet 一般公式的

重要证明，等等。最早的示性类虽由 E. Stiefel 与 H. Whitney 在 1935 年时分别循不同途径引入，但性质所知不多且未定名，直到后来才定名为 Stiefel－Whitney 示性类。由于这些类都是模 2 系数的同调类，因而对微分几何与分析的研究作用有很大局限性。至于整系数的 Pontrjagin 示性类则虽已在 1944 年为 Pontrjagin 所引入，但也未定名并鲜为人知，而且它们的性质直到现在还有很大的神秘性。因而当陈师在 1943 年初次抵美时，纤维丛理论还在萌芽阶段，示性类的概念也处于模糊的状态。但在陈师抵美后的短短几年间，由于陈师的几篇历史性的名著而使纤维丛与示性类理论整个地为之改观。在陈师的 “Characteristic classes of Hermitian manifolds” 一文中，引入了后来被称为陈类的示性类，并提出了多种不同形式的定义。以后的研究证明 Pontrjagin 示性类可以经流形或纤维丛的复化作为陈类来处理，因而陈类在各种示性类中可以说是最基本最有应用前景的一类。后来的发展完全证实了这一点。它们不仅是微分拓扑、微分几何、复流形理论、代数几何等许多不同领域的研究所不可缺少的有力工具，并是使这些不同领域融合在一起的纽带。最后十几年的研究还指出了陈类与 Yang－Mills 场以及其它物理问题有密切关系，因此连理论物理学家们对于陈类这一名称也已耳熟能详，甚至使用到他们的理论物理研究中去了。

凡事必须从根本做起。大范围微分几何的真正发展一直要等到代数拓扑和李群为它铺平道路。因而，尽管陈师的主要目标是大范围微分几何，但在中研院数学所的三年期间，对年轻人没有讲授微分几何，而致力于代数拓扑方面的培养。陈师并对我们这些年轻人指出，要进入近代数学之门，应该好好学习三本书：Pontrjagin 的连续群论，C. Chevalley 的李群论，以及 H. Weyl 的古典群论。事实上，正如陈师早在 40 年代所证

明并在 60 年代为 M. F. Atiyah, R. Bott 等所继续的那样，示性类可以作为某些古典李群作用在纤维丛时的不变量，并由此可以导出它们的明显表达式。

70 年代以来，陈师经常前来中国。多年来作过不少演讲也开过不少课程，但内容都是微分几何。由陈师倡导举办了多次的双微会议，也以微分几何与微分方程为主题。这期间很少讲代数拓扑或微分拓扑。事实上，中研院数学所的三年，陈师已为我国培养了一批拓扑学的骨干，而且代数拓扑除留下一些难题如 Poincaré 推测等外，已非当年之居于数学发展中心者可比。与之相反，国内对 E. Cartan 的著作仍然陌生，对于大范围微分几何更近于空白。陈师这些年来倡导双微，并经常以演讲与课程形式，培养青年一代掌握现代微分几何的要领。如果把国内现在的形势与 70 年代初期相比，则可看出，中国已涌现了一批现代微分几何的少壮队伍，在某些课题方面，已经可使国外专家们刮目相看，取得了一定的国际地位，这是与陈师这些年来的辛勤耕耘分不开的。南开数学所更是有计划地逐年以数学的某些特定范围为中心，邀请外籍专家以及国内有成就的数学家来所系统讲学，鼓励国内青年学者来所进修，已形成一个中外瞩目的国际数学中心。当年中研院的数学所，已以更大更新的规模重见于今日。

陈师一直关心中国数学发展的前途，也一直为促使中国未来成为数学大国而努力。先后两次的数学所，都具有同样的目的。陈师在“二十一世纪中国数学展望”学术讨论会开幕式上提出[1)]：“中国数学的目的是要求中国数学的平等和独立。中国的数学要能够跟西洋的数学平等”，又说：“我们也要求独立，就是说，中国数学不一定跟西洋数学做同一方向，但是要有同样的水平。”为了达到这一目的，必须“在中国建立基地”。两次数学所之设，也正是这方面的具体措施。陈师把这

方面的成功特别寄托在青年一代身上。在中研院数学所，陈师主要是找一些青年人传授现代数学，特别是拓扑学。尽管时间短暂但已经取得极大成功。南开的数学所以及陈师倡导或亲身实行的许多其他活动，也是以提高青年人的学术水平进入研究创作为目的。作为中华民族的优秀青年，如何实现这一宏伟目标，使中国的数学能达到平等和独立，并进而在21世纪使中国成为数学大国，应该是在此书鼓舞之下的一项神圣使命。

《吴文俊文集》前言*

本书收集了作者多年来发表的一些零星文章，散见于各种书刊。这里专门的创作论文收集的很少，因而严格说来，本书是一本杂文集，而不是通常数学家们以专门论著为主要内容的选集或全集。虽然如此，作者认为，这样的文集，至少对作者本人来说，要比出一本或几本选集或全集，其意义要重要的多。因为它真正反映了作者对整个数学的认识，反映了作者思想的实质，也反映了作者对发展数学的主张和意图，而这是过于专门的选集或全集所难以做到的。

本书收集的文章分成四类。但除个别文章外，总的说来是一个整体。基本上它以数学机械化的思想贯通全书。第四类直接以数学的机械化为其主题。从数学有史料为依据的几千年发展过程来看，以公理化思想为主的演绎倾向，以及以机械化思想为主的算法倾向互为消长。对近代数学起着决定作用的解析几何与微积分，实质上都是机械化思想而非公理化思想的产物。中国古代数学，乃是机械化体系的代表。第一类数学史，多少说明了数学机械化思想的渊源，也提供了某些具体的实例。第二类数学论证，则从不同的角度阐述了作者对数学机械化的想法。至于第三类数学专论，表面看来似系专题论著而与机械化无关。但是，推行机械化的前提是数学必须以构造性的方式进行。这一类中主要的四篇文章就是以构造性的方式来讨

---

* 本文摘自《吴文俊文集》，山东教育出版社，1986年。

论这些专题的。这与国外同样的专题的多数论著有着本质的区别。这些文章也多少指出，通常以非构造性方式进行研究的某些数学领域，如果从构造性的角度出发，是有可能趋近于半机械化以致机械化并促使其实现的。

作为数学两种主流的公理化思想与机械化思想，对数学的发展都曾起过巨大的作用，理应兼收并蓄，不可有所偏废。但是对于个别人来说，难免有不同的倾向。作者的倾向如何，应是不言而喻的。但是作者关于机械化思想的形成，决非一朝一夕，至少在70年代以前，机械化的概念在作者脑海里还毫无踪影。经过对中国古代数学的学习的触发，结合着几十年来在数学研究道路上探索实践的回顾与分析，终于形成了这种数学机械化的思想。这种思想一旦形成，就自然地化成一股顽强的动力。十几年来，作者一直在这一方向道路上摸索前进，艰苦奋斗，义无反顾。本书收集的绝大部分文章，都写在这一时期，即是摸索过程中的产物。

我们的目标是明确的，即是推行数学的机械化，使作为中国古代数学传统的机械化思想，光芒普照于整个数学的各个角落。自然，我们离目标的实现还无比遥远根本看不到头。虽然如此，十几年来的努力尽管成就不多，但也已初见成效，至少已可说明这一机械化的道路不仅是可取的，也是可行的。问题不在于能不能成不成，而在于愿不愿做不做，也在于肯不肯敢不敢。

个人的力量是渺小的。事业的成功与否，除了客观条件以外，主要依赖于群众的支持。众志成城，也只有众志才能成城。如果这本书能引起某些有志者的共鸣，乐意共襄此举，作者本人则大为欣慰，本书的出版也就不是多此一举了。

慎重地改革数学教育*

说到数学教育改革，我首先想到的是要慎重，不可草率从事。这是因为它牵涉面很大：全国有两亿中小学生，他们的数学素质如何，直接会影响未来世纪的国家建设。我们的本意是想改得好一些，但是搞得太急，没有充分论证和经过试验，那是不成的。

数学家谈数学教育改革，不能只从培养数学家的角度来看问题。一万人口中顶多有一、两个数学家，不能用数学家的要求来指导中小学数学教学。我们常常以自己如何走上数学道路的经验来判断是非，那是不全面的。

我比较喜欢几何。这里先谈谈中学几何课程的改革。我常常听到一些意见，认为中学的几何必须是一个公理系统。我不赞成。中国古代的几何学，没有公理体系，但是有原理，例如出入相补原理等等。中学几何课上，讲公理不如讲原理。例如三角形全等的条件，就是一个原理。我们选择若干个原理，将几何内容串起来，比公理系统要好。一部经典力学，就是从牛顿三大定律（三个原理推演出来的。也有人认为，从原理出发不严格，使用公理体系才能做到严密。这是在唬人、骗人，中学几何课程根本做不到希尔伯特《几何基础》那样的严格性。欧几里得《几何原本》里的公理体系也是不严格的。我们没有必要去追求这种公理系统的严密性。

---

* 本文摘自《数学教学》，华东师范大学，1993 年 5 月。

当然，我决不是否认逻辑推理的重要性。一旦把几个重要的原理确定下来，我们还是要一步一步地严格论证，从原理出发，推出那些几何学命题和结论。另一方面，几何学有形象化好处，几何会给人以数学直觉。不能把几何学等同于逻辑推理。应该训练学生的逻辑推理能力，但也应适可而止。只会推理，缺乏数学直觉，是不会有创造性的。

不论是几何，还是代数，都要讲推理。你在解方程时，把一个方程化成另一个方程，就要讲“同解”的道理。使用一种算法解问题，也要论证其合理性。任何数学都要讲究逻辑推理，但这只是问题的一方面。更重要的是用数学去解决问题，解决日常生活中，其它学科中出现的数学问题。学校里给的数学题目都是有答案的。已知什么，求证什么，都是清楚的，题目也一定是做得出的。但是将来到了社会上，所面对的问题大多是预先不知道答案的，甚至不知道是否会有答案。这就要培养学生的创造能力，学会处理各种实际数学问题的方法。但要做到这一点，光凭逻辑推理是不够的。

顺便说一句，中学里“对数”的地位似乎应当重新估计。过去学对数主要是为了查对数表以便简化计算。现在有了计算器，对数的这一功能已被取代了。至于对数函数，那恐怕还是要的。

大家谈到我的研究工作，即数学定理的机械化证明，是否可用于中学数学课程改革，我也没有把握。我只是在一个会上谈了设想，有些同志觉得可以试试。至于是否可行，现在还不知道。由于电子计算机的出现，解析几何的重要性在增加，中学里结合解析几何方法学习平面几何，也许值得作进一步研究。不过，我仍回到我刚开始说的一句话，即数学教育改革一定要慎重考虑。一定要经过试验，而且首先要在教师中进行试验。中学里渗透机械化证明的思想，也一定要慎重才好。

在《中国现代数学家传》首卷出版座谈会上的讲话*

我很荣幸收到请柬参加今天江苏教育出版社与中国数学界的“结婚典礼”（因先生手持的请柬乃婚礼专用聘帖，故风趣地称此次会议为“结婚典礼”——整理者注）。刚才程民德教授代表数学界，周（肇锡）、何（振帮）两位分别代表编辑部与出版社介绍了两方面的“恋爱经过”。我想这件事是值得祝贺的。

这本《中国现代数学家传》就象程民德主编刚讲的那样，有“寓史于传”的特色。本世纪可以说是中国数学新的时代的开始。而中国数学进入新的阶段，经历了很长的一段历程，“寓史于传”就可以了解整个中国数学这段发展的历史，要做到“寓史于传”就要考虑许多数学家的传记。现在已经有好几本这一类的书出来了，可是一写传，往往着重于一些对数学有创造性研究、有突出贡献这样的人物。我感觉光写这些人，恐怕不足以反映整个中国现代数学发展的历程。当然，像陈省身、华罗庚这样的许多杰出的数学家，是应该写他们的传，但光写他们我说是不足以反映全貌的。譬如陈省身他自己就经常提到姜立夫、孙塘对他成长的重要影响。那么，提到华罗庚恐怕就不能不提熊庆来、杨武之、唐培经等人。而熊庆来、杨武之、唐培经、姜立夫、孙塘等人又有他们自己的成长经历。所

---

* 本文摘自《中国现代数学家传》第二卷，江苏教育出版社，1995年12月。

以，要考虑这些人就不能不考虑他们的背景与历史条件，而要整个地把这些都反映出来，就不能光写创造性成就。我觉得这本书好的地方、突出的地方就是它包括了除那些在研究上有突出贡献的人物外，还介绍了20世纪整个数学发展中，用各式各样的方式有所贡献的人物。譬如程民德先生刚才提到的这本书第一个入传人吴在渊，如果仅局限于研究的话，我想吴在渊的工作是不值得提的，但是他对中国数学发展有很大作用。我是在上海出生的，在我年青的时候，吴在渊的名字在上海数学界无人不知、无人不晓。我自己也看了吴在渊的许多文章，主要是关于初等几何方面。如他介绍的若三角形两内角的平分线是相等的，则它应是等腰的，这是很难证的题目，他写文章介绍很多种不同的证法。还有其他很多问题，具体内容我已记不清了。这对我个人有一定启发引导作用。我个人感觉，在当时上海数学界，能够领导许多青年人走上数学道路，吴在渊起到了很大作用。像这样的人，假定只着眼于数学研究，在数学上有很大贡献这个角度，这些人就会被忽略掉，但要了解整个中国在20世纪现代数学的发展过程，这样人又是不能不考虑在内的。在目前各种传记，包括数学史、数学家传中，被忽略掉的人物在这本书里就可以找到。我所举的仅是一个例子。因吴在渊是我比较了解的一个，这样的人物还有不少。

我走上数学道路是受了上大学时的一位教授的影响。我进入大学数学系时本来已不想学数学了，是受这位教授的影响才决定学数学的。但要写我的传记可能还不会提到这位教授，而要提起陈省身的，这种情形恐怕是普遍的，但要真正了解历史，就要把这些人都包括进去。当然这是不容易的事。如要选哪些人入传，又要找了解他的合适的人给他写传。一个人了解的可能有限，还要为其提供资料，要有许多人来提供资料。编辑部与出版社花了很大的力气来组织这件事，是非常不容易

的！

我希望这本书只是这个婚姻的初步结晶，我因此祝贺这个婚姻圆满成功！

法国数学新派——布尔巴基派*

近 20 年来法国有一部分青年数学家以 N. Bourbaki 为名，兴起了对数学的一种革新运动，数学发展到了 20 世纪分支愈加复杂，会有人认为数学已划分为许多不同的畛域，各有各的特点和界限，仅有少数路径可以互相沟通学者们终其一生，只能在一隅之地作狭而深的研究，要懂得全部数学已不可能，但 Bourbaki 却抱着极大野心想用统一的方法和统一的观点冶数学全部於一炉。他们认为到了目前，数学在表面上虽然部门增加，方向繁多，事实上却比以前更加统一。因此法文的数学原名 Les mathématiques（多数），Bourbaki 派把它改成 la mathé matique（单数）。

为此 Bourbaki 派创造了“构造”（Strutures）一词，统一了数学研究的对象。所谓构造，可以说是表示一个集合中各元素之间的关系而把他们组织起来的一种方式。试举一切实数所成的集合 R 为例，在 R 的各元素——实数——之间存在着下面三种关系：1°实数可按大小排列；2°任两实数可以相加相乘以得另一实数；3°一串实数有时有极限值把这种关系抽象化，可能得到集合的三种构造。

1°序次构造　两个元素间可有某种关系＞满足下列条件：

如 $a>b$ 及 $b>c$，则 $a>c$。

2°代数构造　有一种或数种满足适当条件的结合（或运

---

* 本文摘自《数学通报》，1994 年第 4 期。

算）方法可从两元素得一第三元素，视所加的条件不同可得不同的代数构造如群（Groupes）环（anneau）体（Corps）等，上面所说的实数集合 R 是一个“体”。

3°拓扑构造　在原来的集合中有满足适当条件的一组部分集合，由此可导出极限和连续等观念。有了拓扑构造的集合称为空间（espace）。

上面屡次提到的“条件”一词，在数学上称为公理（Axiomes）。

在 Bourbaki 派的分析之下，数学无非是许多简单与复杂，普遍与特殊的种种构造的研究。上面所说的三种构造可以说是数学的“基本构造”。在一个集合里面同时讨论几种不同的基本构造，用若干公理把它们联系起来，则可得到比较复杂的“联合构造”。例如实数集合 R 即是序次，代数，拓扑三位一体的一种联合构造。Lattice 可以看做既有序次又有∪、∧两种结合方法的一种序次和代数的联合构造，研究这一类构造的数学部门就叫造 Lattice theory。同样，研究代数与拓扑的联合构造者叫拓扑代数学。在一个空间的某种部分集合间定义种种结合法则而研究之者叫代数拓扑学。若把所讨论的集合加以明确的规定，则我们又可得到特殊的数学部门，如实变数或复变数函数论等，那时候的集合是实数集或复数集，不再是任意的集合了。总之，数学建筑就如一座城市以三种基本构造为中心，以各种联合构造为郊外。它的中心时时在重建与改造，它的郊外则不断膨胀与扩展。

Bourbaki 派依照他们的观点计划写一部大书统括数学全部。这部书的第一篇讨论分析学的基本构造拟分六卷。卷一论集合和构造的意义兼及序次构造，全书计一册。卷二论代数构造已出四册。卷三论拓扑构造已全出，计五册，卷四论一个变数的实变数函数，已出一册。卷五论积分，卷六编 Espaces Vec-

toriels topologiques，都未出书。这一部份在三四年中可以出全。那时候 Bourbaki 派理想中的数学城的中心部分可以说是落成了。这部书的其他部分拟讨论数学的上层建筑，内容和篇数却尚无具体计划。全部写成至少尚需 20 年，可能要写 50 年。

Bourbaki 派每年在巴黎举行公开的演讲会三次，又在 Nancy 和 Strastourg 每年举行不公开的集会若干次，讨论书籍的编写问题。

数 学 史

中国古代数学对世界文化的伟大贡献*

自秦至西汉中期这两百来年间，我国工农业生产与科学技术有了很大的进步，从而促进了数学的迅猛发展，出现了一批高水平的数学家，如张苍、耿寿昌等。《周髀算经》、《许商算术》与《杜忠算术》(后二者已失传)，都在这时期出现。我国最主要的一部传于后世的数学著作《九章算术》，也基本上成书于西汉初年，其内容为以后一千多年的辉煌成就奠定了基础。从西汉以讫宋元，随着我国社会经济和劳动人民创造性发展，数学人才与数学创作仍世代不绝，中国的数学，在世界上可以说一直居于主导地位并在许多主要的领域内遥遥领先，直到宋末明初，宋明理学成为垄断一切的统治思想，明代并以八股取士，以及其他一些原因，科学技术的发展受到扼杀，除了民间的计算技术还有重要发展外，数学已相应地大为衰落。从明末利玛窦怀着不良企图以介绍西方数学为名打入我国统治集团内部以来，我国数学与古代相比已谈不上什么创造，基本上依靠国外的技术输入，在外国人屁股后面爬行了。正如毛主席批评的那样“言必称希腊，对于自己的祖宗，则对不住，忘记了。”西方的大多数数学史家，除了言必称希腊以外，对于东方的数学，则歪曲历史，制造了不少巴比伦神话与印度神话，把中国数学的辉煌成就尽量贬低，甚至视而不见，一笔抹煞。在半封建半殖民地社会中生活过来的一些旧知识分子，接触的

---

* 本文摘自《数学学报》，1975年，第18期。

数学都是“西方”的，看到的数学史都是“西方史家”的，对于祖国古代数学十分无知，因而对于西方数学史家的一些捏造与歪曲无从辨别，不是跟着言必称希腊，就只好不吭声。

但是，被颠倒了的历史必须颠倒过来！

中国古代数学的成就在本世纪五六十年代，或更早的中外中小学数学教本中，可找到许多具体的例子。

有一本西方数学史[12]，（这基本上是一本比较好的数学史）说：“在许多中学中，代数学现仍被教成一堆公式而不是一种演绎的科学。”

在谈到代数学的东方起源时又说：

“今日学校中的代数学和几何学仍然保持这些不同来源的标志”。

这里所谓代数学的东方起源或东方数学，乃是针对“由一些定义、假定和公理到定理的一种严格逻辑的演绎法”作为处理问题准则的西方（或希腊）式数学来说的。这本书所说的东方，原意是指巴比伦或是印度，而事实上这个东方应指中国才算确当。

中小学数学中的算术、代数这些部分，从记数、以至解联立线性方程与二次方程，实质上都是中国古代数学家的发明创造，早就见之于中国的《九章算术》甚至是《周髀算经》等书。据钱宝琮考证（[3]），《九章算术》完成于公元 50—100 年间。但除个别片段外，基本内容应完成于公元前 200 年或更前一些（这是某些西方数学史家的意见。有的甚至提早到公元前 1000 年，例如[9]）。另一部《周髀算经》据钱宝琮考证（[3]）成书于公元前 100 年前后。

下面是关于算术代数部分发明创造的一张中外对照表，这里应该指出，表中虽列有印度的发明，但诚如一位印度数学史专家 Kaye 所说的那样：印度与中国的数学有很多平行之处，

	中　　国	外　　国
位值制十进位记数法	最迟在《九章算术》成书时已十分成熟	印度最早在6世纪末才出现
分数运算	《周髀算经》中已有，在《九章算术》成书时已成熟	印度最早在7世纪才应用
十进位小数	刘徽注中引入，宋秦九韶1247时已通行	西欧16世纪时始有之，印度无
开平方、立方	《周髀算经》中已有开平方，《九章算术》中开平、立方已成熟	西方在4世纪末始有开平方，但还无开立方 印度最早在7世纪
算术应用	《九章算术》中有各种类型的应用问题	印度7世纪后的数学书中有某些与中国类似的问题与方法
正负数	《九章算术》中已成熟	印度最早见于7世纪，西欧至16世纪始有之。 所谓公元3—4世纪Diophantus有正负数规则之说是有问题的
联立一次方程组	《九章算术》中已成熟	印度7世纪后开始有一些特殊类型的方程组 西方迟至16世纪始有之
二次方程	《九章算术》中已隐含了求数值解法 三国时有一般解求法	印度在7世纪后 阿拉伯在9世纪有一般解求法
三次方程	唐初（公元7世纪初）有列方程法、求数值解已成熟	西欧至16世纪有一般解求法、阿拉伯10世纪有几何解
高次方程	宋时（12—13世纪）已有数值解法	西欧至19世纪初始有同样方法
联立高次方程组与消元法	元时（14世纪初）已有之	西欧甚迟，估计在19世纪

而印度是欠了中国的债（参阅例如 Cajori,[6]，页 97 与 84，又如 Scott[9]）。

中国劳动人民，在长期的实践过程中，创造与发展了记数、分数、小数、正负数以及无限逼近任一实数的方法，实质上达到了整个实数系统的完成。特别是自古就有了完美的 10 进位位值制的记数法。这是中国的独特创造，是世界其他古代民族都没有的。这一创造对世界文化贡献之大，如果不能与火的发明相比，也是可以与火药、指南针、印刷术一类发明相媲美的。

代数学无可争辩地是中国的创造，这从《九章算术》等书中可以看出。可以说在 16 世纪以前，除了阿拉伯某些著作之外，代数学基本上是中国一手包办了。但中国古代数学的成就决不止于算术与代数方面，以几何而论，希腊欧几里得几何的拱心石是毕达哥拉斯定理（语出 Bourbaki,[5]）或勾股定理。这一定理我国古代自然也早已有之。在《周髀算经》中就已有一般定理的叙述：

“若求邪至日者，以日下为句*，日高为股，句、股各自乘，并而开方除之，得邪至日。”

不仅如此，勾股定理还被具体用于勾股弦的直接互求，甚至应用于测日之高远这一类复杂问题。这与欧几里得几何中理论脱离实际的情况是迥不相同的。中国的几何学与希腊的几何学有许多不同之处，其详细比较有待阐发。

对于三角学中国也是最早发明者之一。西方数学史家一般都把《天文书》（Almagest）的作者托雷米（Ptolemy，公元 150 年左右）作为三角术的创始人，而把中国的三角术视为是受了他的影响。

---

* 勾股的勾字，古代均作句，本书引用的原文均用句。

但事实上，西方的三角术是先有球面三角后有平面三角。托雷米的三角术由测天而来，因而是球面三角术。至于平面三角术则迟至公元 1250 年才由波斯天文学家纳速剌丁所建立。但我国一开始建立的就是平面三角术，并且中国的三角术来源于《周髀算经》。东西方三角术的发展途径是刚巧相反的，很难谈到有什么相互影响。如果说有影响，那么《周髀算经》早于《天文书》有好几百年，只有说托雷米受到中国的影响才更合情理。上面所引西方数学史家关于中国三角学之说显然是颠倒了历史事实。

在西欧，16 世纪中出现了不少描述三角测量的图画，其中有一张名为"鼓皮三角法"，所画犹如周髀赵爽注所附的日高图（也即重差原理图）的一个翻版，这也可以见到我国三角测量术的先进程度。(画见 Smith,[10]，页 355)

到西欧 17 世纪以后才出现的解析几何与微积分，乃是通向所谓近代数学的主要的两大创造，一般认为这些创造纯粹是西欧数学的成就。但是中国的古代数学决不是不起着重大作用（甚或还是决定性的作用）的。

先说解析几何，Smith（[11]，页 316,）曾认为解析几何的发展有三个主要阶段：(1) 坐标系统的发明；(2) 几何与代数间一一对应的认识；(3) 函数 $y=f(x)$ 的图形表示。第一阶段属于古代，第二阶段属于中世纪，第三阶段则是近代的。

西方向来认为 17 世纪的 Descartes（以及 Fermat）是解析几何的创始人，但实际上在 Descartes 的有关主要著作中既无坐标也无坐标轴的概念，更无所谓直线与曲线的方程。Descartes 的贡献在几何与代数建立关系方面，在他的主要著作中，给出了二次方程的几何解法，但阿拉伯最早的代数学即 Al Khowarizmi（花剌子模）的著作（公元 9 世纪）也早已用

另一种较 Descartes 更好的方法给出了二次方程的几何解。事实上，几何与代数的统一处理乃是我国古代数学的一个传统特色，从《九章算术》以来就向来如此。花剌子模的著作据 Cajori[6]与希腊印度无关，如果不是阿拉伯自己的发明创造，则必然渊源于中国，从著作的风格看来，后者是不无可能的。这一段历史自然是值得重视并予以澄清的。现已知花剌子模在842—847 曾出使波斯以北当东西方商业要冲的西突厥可萨国，而可萨通中国语，朝廷依中国礼仪（[13]，Addenda)，详情有待进一步调查。

至于第一阶段，西方数学史家比较一致地认为真正的坐标概念出现于 14 世纪中叶 Oresme 关于以“经度”、“纬度”来表示点的位置的一个著作。据 Smith 指出（[11]，页 320 注)，Oresme 的著作可能导源于 10 世纪时的一个作品。这里 10 世纪的作品估计应是阿拉伯的。在我国，《周髀算经》中已有“分度以定则正督经纬”以及“游仪所至之尺为度数”等语，注中并屡言“引绳至经纬之交，以望之。”中国又有世界上最早的星表（甘石星经，战国中叶，公元前三百五六十年)，公元 2 世纪张衡就作星图与浑天仪，又有世界上最早的石刻星图(宋，公元 1247 年，在苏州)。我国的天文数学历来紧密结合。由此可以看到以经纬度表星的位置的这种坐标概念我国是最早的创始人之一。我国又是罗盘的发明者并曾经是航海最发达的国家之一。用经纬度表位置的概念与方法在后来必然有所发展，其演变以及与阿拉伯西欧的关系，是值得把它追查清楚的。

微积分，这是使西欧数学一跃而居世界领导地位的重大发明创造，我国似乎是没有份的。但是微积分的发明从 Kepler 到牛顿有一段艰难的过程。在作为产生微积分所必要的准备条件中，有些是在我国早已有之，而为希腊式的数学所力所不及

的。例如（见 Scott,[9]，页 138）：

“极限的概念，作为微分学的真正基础，对于希腊头脑来说完全象是一个外国人。”

希腊数学中被认为最辉煌的创造之一的无理数论，对于极限来说是华而不实的。而刘徽以至宋代的我国十进位小数的记数法，却与极限概念一衣带水。十进位小数迟至 16 世纪在西欧重被发明以来，直接导致了对数的发明。作为微积分先驱者之一的 Kepler，“广泛应用了对数与十进位分数，且热情地传播这方面的知识。”（Cajori,[6]，页 160），是有一定的道理的。面积体积的计算乃是导致微积分发明的另一重要问题。然而，原来希腊欧几里得以至阿基米德所使用的“穷竭法”是很不得力的，Kepler 用之劳而少功，直到伽利略的学生 Cavalieri 放弃了严密的穷竭法改用粗糙的不可分量法才取得了重大的突破。在微积分的创造过程中起了如此重大作用为西方数学史家盛称的所谓 Cavalieri 原理，事实上早就见之于祖冲之、祖暅父子的著作，即所谓“幂势既同则积不容异”并具体用之于球体积的计算，比 Cavarieli 的发现要早了 1100 多年。

微积分的发明从 Kepler 与 Calileo 以至 Newton 与 Leibniz 经历过一段艰苦漫长的过程。上面所举两个例子可以说明发明过程中中国古代数学的作用远优于希腊式的数学，我们甚至不无理由可以这么说，微积分的发明乃是中国式数学战胜了希腊式数学的产物。

中国古代数学至少自秦汉有记载以来，许多方面一直居于世界上遥遥领先的地位，发展到宋元之世，已经具备了西欧 17 世纪发明微积分前夕的许多条件，不妨说我们已经接近了微积分的大门。如果按此继续发展下去，我们是有可能先于欧洲发明微积分的。然而，宋朝的程朱理学已使当时的一些优秀数学家（例如杨辉）浪费精力于纵横图之类的数学游戏，陷入

神秘主义，违反了我国自古以来的优良传统，到了明朝八股取士，理学统治了学术界的思想，我国的数学也就从此一落千丈了。

西方数学史家往往以希腊式的严密推理相标榜，并以中国数学从来没有达到演绎科学的形式相指责。然而，我们已经看到，在微积分的发明上希腊形式的那种脆弱性以及中国式数学的生命力。某些数学史家例如 Bourbaki[15]也曾指出欧几里得的那种系统阻碍了代数学的发展并使之瘫痪。在将 Cavalieri 与阿基米德作比较时，Bourbaki 又指出阿基米德只能得到 Cavalieri 原理很特殊的情况，而与 Cavalieri 作出他的原理用了不很科学的所谓“证明”相仿，阿基米德为了获得他的特殊情况的“证明”，也不得不把他著名的所谓阿基米德严密性弃之脑后。我国古代数学并没有发展出一套演绎推理的形式系统，但却另有一套更有生命力的系统。刘徽《九章注》序中说“析理以辞，解体用图”。刘徽《海岛算经》本来有注有图，注以析理，图以解体，只是已失传而已，这是古代数学用以分析矛盾解决矛盾的一种辩证思维方法。中国古代的劳动人民向来重视实际，善于从实际中发现问题提炼问题，进而分析问题解决问题，在深入广泛实践的基础上建立了世界上最先进的我国古代数学。中国的数学是牢牢扎根于广大劳动人民之中，是导源于劳动人民长期实践经验的基础之上的，这与希腊几何学脱离实际脱离群众走到纯逻辑推理的形式主义道路是有别的。这正是直至 16 世纪以前我国数学在许多最主要的领域内一直居于最先进地位的根本原因，也是在微积分的发明上中国式的数学远远优越于希腊式数学的根本原因。西方数学史家把它归之于我国数学的缺少演绎推理与历史事实完全不符。恩格斯曾经说过(见[1])：

“数学演算适合于物质的证明，适合于检验，因为它们是

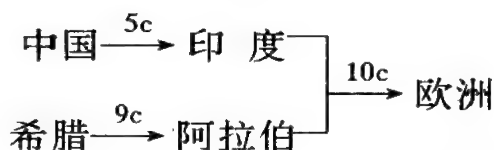
建立在物质直观（尽管是抽象的）的基础上的；而纯逻辑演算只适合于推理证明，因此没有数学演算所具有的实证的可靠性——而且其中许多还是错误的！”

这是我国数学对希腊式数学来说具有极大优越性的一个很好的说明。

钱宝琮在《中国古代数学的伟大成就》一文[4]）中曾说：

“第5世纪以后，大部分印度数学是中国式的，第9世纪以后，大部分阿拉伯数学是希腊式的，到第10世纪中这两派数学合流，通过非洲北部与西班牙的回教徒，传到欧洲各地，于是欧洲人一方面恢复已经失去的希腊数学，一方面吸收有生力量的中国数学，近代数学才得开始辩证的发展。”

这段数学发展过程可概括为下面的简图（c表示世纪）：



根据前面的论证，我们认为有理由可以进一步说：近代数学之所以能够发展到今天，主要是靠中国的数学，而非希腊的数学，决定数学历史发展进程的主要是中国的数学而非希腊的数学。

以上抛砖引玉，论证粗疏不全之处，希望进一步补充阐发。论证偏激不当之处，则希望引起争鸣。

参 考 文 献

[1] 马克思数学手稿，北京大学学报专刊，1974年5月。

[2] 算经十书（钱宝琮校点），中华书局，1963。

[3] 钱宝琮，中国数学史，科学出版社，1964。

[4] 钱宝琮，中国古代数学的伟大成就，科学通报，1951，2卷，10期，1041—1043。

[5] Bourbaki N., Eléments d'histoires des matheinatiques, 1969.
[6] Cajori F., A history of mathematics. 2nd ed, 1919.
[7] Datta B., Singh A., N., History of Hindu mathematics, 1962.
[8] Mikami Y, (三上义夫), The development of mathematics in China and Japan, 1913.
[9] Scott, A history of mathematics, 1958.
[10] Smith D. E., Source book in mathematics, 1929.
[11] ——, History of mathematics, 1925.
[12] Struik D. J., A concise history of mathematics, 1948.
[13] Needham J., scicnce and Civilization of China, V. 3, 1971.

近年来中国数学史的研究*

一、绪　　论

本文的讨论限于对从远古时代至公元 14 世纪的中国古代数学的研究。近年来，国内外学者对中国古代数学的研究已经蓬勃开展起来，对中国古代数学所取得的成就有了更深刻的认识。笔者在此将随意引用其中一些成果，对文中所表述的观点，概由笔者本人负责。

在中国古代数学史研究中，有两条需要严格遵守的原则，即

原则一：所有研究结论应该在幸存至今的原著基础上得出。

原则二：所有结论应该利用古人当时的知识、辅助工具和惯用的推理方法得出。

本文将引用下列著作：

《九章算术》，成书于公元 1 世纪。

《九章算术注》，刘徽著，成书于公元 263 年。

《海岛算经》，刘徽著，成书于公元 263 年。

《数书九章》，秦九韶著，成书于公元 1247 年。

就原则二来说，因为中国古代数学中没有使用代数符号演算和添加平行线证明几何题的传统，所以我们强调应该在代数

---

* 本文摘自《中国数学史论文集》(三)，山东教育出版社，1987 年。

和几何的推理中禁止使用诸如此类的“现代”方法。中国古代数学独立于古希腊数学和作为其延续的西方数学，有着自身发展的清晰主线，其发展过程、思考方法和表达风格亦与西方数学迥然不同。笔者在开始对中国古代数学成就进行叙述之前，首先在此指出中国古代数学的一些特点。

第一，古代中国人没有用纸笔作为计算工具，而是在算板上用算筹进行一切计算。为了使算筹运算成为可能，早在远古时期，我们祖先已经使用了十分完善的十进位位值制。只要把筹摆在板上的适当位置即可准确表示一个整数，特别地，十进整数中的零和零本身都通过在相应位置上空位表示。事实上，我国古代把数学称作“算术”，正好反映了这种计算的方法。

第二，结果通常是由独立问题的形式来表示。每个问题由以下条目构成：（一）叙述具体数值的问题；（二）问题的数值解答；（三）术文，叙述获得结果所用的方法，在大多数场合相当于现在的“算法”，但有时也相当于现今的公式或定理。值得注意的是，（一）中的具体数值，在术文中并不起特殊作用，以其他数值代入运算也同样成立。因此，术文是一般方法，条目（一）中的具体数值只是作为举例说明之用。（四）有时有“注”或“按”，解释术文所依据的理论。宋代以后的算书中，常会有条目（五）演草，抄录获得最终结果所进行运算的详细步骤。

二、关于整数理论的研究

本部分整数系指正整数。

在中国古代数学中，未曾出现素数、因数分解等概念，但是发明了求两整数的最大公因数的方法——更相减损术。其算法如下：

“以少减多，更相减损，求其等也。”

在中国古代数学中，将最大公因数称作“等”。

举个简单的例子，用更相减损术求得 24 和 15 的最大公因数是 3，其步骤如下：

$$(24,15)\to(9,15)\to(9,6)\to(3,6)\to(3,3)。\qquad (1)$$

正如刘徽在《九章算术注》中所说：“其所以相减者，皆等数之重叠。”此术所依据的原理是在更相减损过程中整数值逐渐变小，而最大公因数不变。

中国古代虽然未出现过素数概念，但却从不同于西方数学的角度对整数理论进行了相当广泛和深入的研究。我们将引用南京大学莫绍揆和西北大学李继闵的两项研究工作。

在中国古代数学发展的漫长过程中，勾股形一直是数学家们研究的偏爱对象。特别是勾、股、弦三边长度所组成的三重整数组已经完全可以确定。例如经典算书《九章算书》“勾股”章中就出现了以下 8 个三重数组：(3，4，5)，(5，12，13)，(7，24，25)，(8，15，17)，(20，21，29)，(20，99，101)，(48，55，73)，(60，91，109)。这些三重数组的出现不是偶然的，在勾股章第十四问中已隐含给出这种数组的一般组成形式。今将此题照录如下：

今有二人同所立，甲行率七，乙行率三。乙东行。甲南行十步而邪东北与乙会。问甲乙行各几何。”

术文为“令七自乘，三亦自乘，并而半之，以为甲斜行率。斜行率减于七自乘，余为南行率。以三乘七为乙东行率。”

如前所述，数字 7 和 3 仅起举例说明的作用，可以用满足 $m>n>0$ 的任意数对 m、n 替换它们。术文所说就是三边成这样的比例：

$$勾:股:弦=\left(m^2-\frac{m^2+n^2}{2}\right):mn:\frac{m^2+n^2}{2}。$$

上述8个三重数组就可以用下数对来决定：$(m, n)=(2,1),(3,2),(4,3),(4,1),(5,2),(10,1)(8,3),(10,3)$。

刘徽在《九章算术注》中，依据出入相补原理对此问作了几何证明。本文第三部分将详细讨论出入相补原理。刘徽的证明表明 $m:n$ 即股弦和与勾之比当且仅当勾股弦的长度成一个整数比时，$m:n$ 是一个整数比。因此，利用术文所给的方法，可以列出由勾股弦长度所组成的详尽数表。

中国古代数学中对整数理论研究的第二个例子是大衍求一术，即著名的中国剩余定理。最近的研究表明大衍求一术算法来源于汉代以来的历法计算，直到公元1247年秦九韶《数书九章》的问世，有一条相当清晰的发展路线。秦九韶在书的序言中指出："独大衍法不载《九章》，未有能推之者，历家演法颇用之。"秦九韶第一次系统地解释了大衍求一术算法。《数书九章》卷一、二共九问，内容有历法计算，修筑河堤，计算财物，租税分配，出售谷物，计点军队，土木建筑，甚至还有一个盗窃案例。如果使用现代符号表示，那么所有问题都可归结为以下形式：

$$u \equiv u_j \pmod{M_j},\ 1 \leqslant j \leqslant r。 \tag{2}$$

其中 u_j，M_j 已知。秦九韶称模 M_j 为定母（不必两两互素），秦给出一种方法，可以反复运用更相减损术把定母化为两两互素，因此这里只考虑模 M_j 两两互素的情况。

在现代数学中这一问题可以这样解决：

令 $Ph_j(N)$ 表示整数 N 的欧拉函数，此函数可以将 N 进行素因数分解得到。

令 $M=M_1 \cdot M_2 \cdot \cdots \cdot M_r$，

$$N_j=\left(\frac{M}{M_j}\right)^{Ph_j(N_j)},\ 1 \leqslant j \leqslant r \tag{3}$$

则（2）的解为

$$u=\sum_j (u_j \cdot N_j) \pmod{M}。$$

读者可以参看 Knuth：Art of Computer Programming 卷 2 第 250 页。

(3) 的方法和结果似乎很简洁漂亮，然而用此法计算秦九韶的 9 个问题，即使有现代计算机辅助，得到最终结果也是相当困难的。

现在，我们介绍秦九韶的解法。

第一步，令 R_j 为$\frac{M}{M_j}$模 M_j 后的余数，称为奇数，决定 K_j，使之满足：

$$K_j \cdot R_j \equiv 1 \pmod{M_j}。\tag{4}$$

最后结果：

$$u \equiv \sum_j \left(u_j \cdot K_j \cdot \frac{M}{M_j}\right) \pmod{M}。\tag{5}$$

K_j 称为乘率，求 K_j 满足 (4) 的算法被秦氏称为大衍求一术。此术的第一步是将四个已知数 1、0（留作空白）、R_j 和 M_j 分别放在正方形左上角 (LU)，左下角 (LL)，右上角 (RU)，右下角 (RL)：

$$\begin{array}{|cc|}LU & RU \\ LL & RL\end{array} = \begin{array}{|cc|}1 & R_j \\ & M_j\end{array}$$

这四个数满足下列同余式：

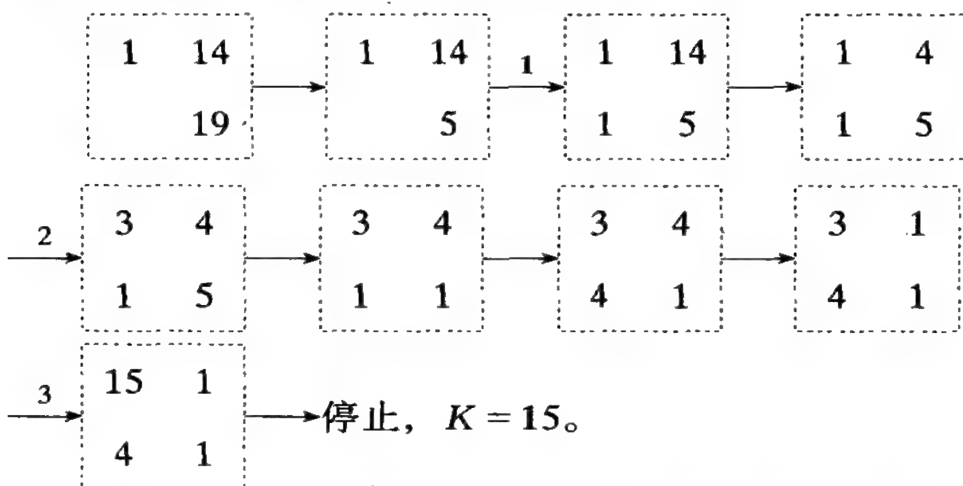
$$LU \cdot R_j \equiv RU \pmod{M_j},$$
$$LL \cdot R_j \equiv -RL \pmod{M_j}。$$

算法第二步是对正方形四个数进行一定的运算，不断减小右方的数值，但保持同余式 (6) 的成立。经过有限次运算后，右上角的数值将减小为 1，由 (6) 知左上角的数值就是所求的乘率 K_j。通过下面对算法的详细叙述，读者不难看出，大

衍求一术所依据的原理本质是与求两数最大公因数的更相减损术相同，只不过大衍求一术比更相减损术复杂得多罢了，求乘率的算法如下：

“以奇为右上，定母为右下，立天元一于左上。先以右行上下两位以少减多，所得商数，乃递互乘内左行，使右上得一而止。左上为乘率。”

今以秦书第九问为例。这是一个盗窃问题。负责此案的官员使用这种算法，就可判断出三个盗贼每人所窃的稻谷数量。计算其中一盗贼所窃物时，其相应乘率：



读者可将这一过程与（1）的计算过程作一番比较，便不难体会大衍求一术与更相减损术的异同之处。

此例的数值是秦书九问中最简单的，但若用（3）的方法计算则不是一桩易事。另有三问中出现了庞大的天文数字，如用（3）的方法最终将鞭长莫及，但秦九韶用大衍求一术轻松地解决了这九个问题。

三、几　　何

通常认为，中国古代没有几何学。事实上却不是这样，中

国古代在几何学上取得了极其辉煌的成就。人们的误解可能是因为中国古代几何学在内容和形式上都与欧几里得几何迥然不同的缘故。这种不同表现在以下几方面：

第一，中国古代几何没有采用定义——公理——定理——证明这种欧氏演绎系统，取公理而代之的是几条简洁明了的原理，在此基础上推导出各种不同的几何结果，刘徽在《九章算术注》中就是这样做的。

第二，中国古代几何与欧氏几何研究的侧重点不同。我们祖先对直线的垂直性感兴趣，而欧氏几何重视平行性的研究。在中国古代数学数千年的发展过程中，勾股形在几何研究中始终占据着主要地位。其次，古代中国人对角缺乏兴趣，重视对距离的研究，而欧氏几何则把角的研究放在重要地位。

第三，中国古代几何学总是与应用问题紧密相联，测量、面积和体积的研究占据了研究的中心地位。

第四，中国古代的几何总是与代数互相渗透，具有几何代数化的特点。中国古代数学的几何代数化在宋元时期达到了顶峰。李约瑟指出，几何代数化是解析几何产生的前奏曲（也是关键的一步)。

我们用下面几例来进一步阐述这些观点。

例 1　日高术

“立两表于洛阳之城，令高八尺。南北各尽平地，同日度其正中之景。以景差为法，表高乘表间为实，实如法而一，所得加表高即日去地也。”即

$$\text{日高}=\frac{\text{表高}\times\text{表面}}{\text{影差}}+\text{表高。}$$

在汉代的一些著作中已出现了这个公式，后来的历法书中也常引用它计算日高。显然，这个公式估值过于粗糙而实际上不能使用。刘徽用海岛代替日高术中的太阳，将日高术改造成

为海岛公式，成为一个地面测量实际可行的精确公式。《海岛算经》包含9个类似公式，上面这个公式是第一个，也是最简单的。古代对这个古老公式曾有证明及附图，宋代书中记载了这一事实，以后便失传了。笔者根据3世纪赵爽所作的证明片断和残缺不全的彩图，按照逻辑顺序重排了语句，作出了日高术及海岛公式的证明。图1中 Y 代表黄色，B 代表青色。

“黄甲（Y_1）与黄乙（Y_2）其实相等。黄甲与青丙（B_3）相连，黄与青已（B_6）相连，其实亦等。青丙与青已其实亦等。以表高乘表相与为黄甲之实。以影差为黄乙之广而一，所得则变为黄乙之袤。上与日齐，按图当加表高。”

参看附图，证明是很清楚的。

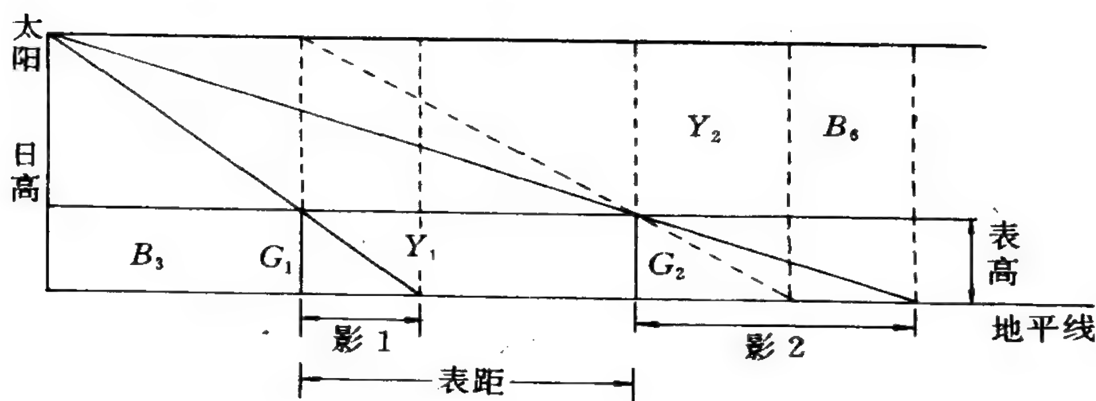


图 1

例 2 出入相补原理

例1中的面积相等关系是运用出入相补原理得出的。在经典著作《九章算术》中可以看出出入相补原理已经形成，并用简洁的语言清楚地表达出来。出入相补原理的含义为，一个平面或立体的图形被分割成几部分后，面积或体积的总和保持不变。这个简单的原理被古人成功地运用，解决了各种各样的复杂问题，有时问题解决的十分巧妙，例1仅是其中一例。仍以勾股形的三边为例。这三边可以形成几种和与差，例如勾股

和、股弦差等等，连同三边有 9 个数。在《九章算术》勾股章中，有许多问题是已知这 9 个数中的两个利用出入相补原理求勾、股、弦。如第十四问已知勾弦和与股的比值，可得出整勾股数的一般公式，如同第二部分中已给出的公式。刘徽用出入相补原理证明了这一结果，如图 2 所示（R 表示红色，Y 表示黄色）：

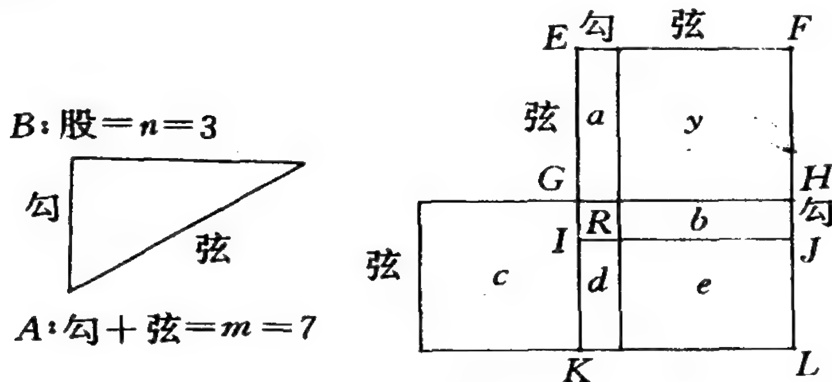


图 2

$$c+d=弦^2-勾^2=d+e=股^2=n^2,$$

$$2EFGH=EFKL=m^2+n^2=(勾+弦)^2+股^2,$$

$$a+Y=弦\cdot m=弦\cdot(勾+弦),$$

$$b+R=EFIJ-EFGH=勾\cdot m=勾\cdot(勾+弦)$$

$$=m^2-\frac{m^2+n^2}{2}。$$

在《数书九章》中有一个已知三角形三边求三角形面积的公式。

$$4\cdot三角形面积^2=短边^2\cdot长边^2-\left(\frac{长边^2+短边^2-中边^2}{2}\right)^2。$$

这公式显然与海伦公式等价。以公式的繁琐与海伦公式的简洁比较，肯定此公式不是由海伦公式得出来的。笔者按照中国古代数学的传统风格，应用《九章算术》中基于出入相补原理的一些公式，重新构造了一个证明，可以很自然地得到秦九

韶的公式。

出入相补原理事实上也是中国古代求平方根或立方根方法和二次方程解法的基础，因而中国古代代数的本质是具有几何性的。用出入相补原理可以很自然地得出《海岛算经》中的所有复杂公式，但是如果用欧氏几何的方法推导这些公式就很困难，至少会使人有不自然的感觉。

例 3 立体体积

运用出入相补原理可以决定多边形的面积，但不能决定所有多面体的体积。刘徽意识到这一点，并出色地解决了这个问题。斜解正方体成两个相等部分，叫做堑堵。斜解堑堵成两部分，其中锥形部分叫阳马，另一个四面体叫鳖臑。刘徽运用极限以清晰的逻辑推理过程，得出了结论，笔者将其命名为刘徽原理。

“阳马居二，鳖臑居一，不易之率也。”

用刘徽原理和出入相补原理，可以求出任何多面体的体积。《九章算术》商功章的许多公式便是这样得出的。在刘徽对其原理的简洁严密的证明中，有将一个大堑堵分割为较小的阳马的过程。如下图 3 所示：

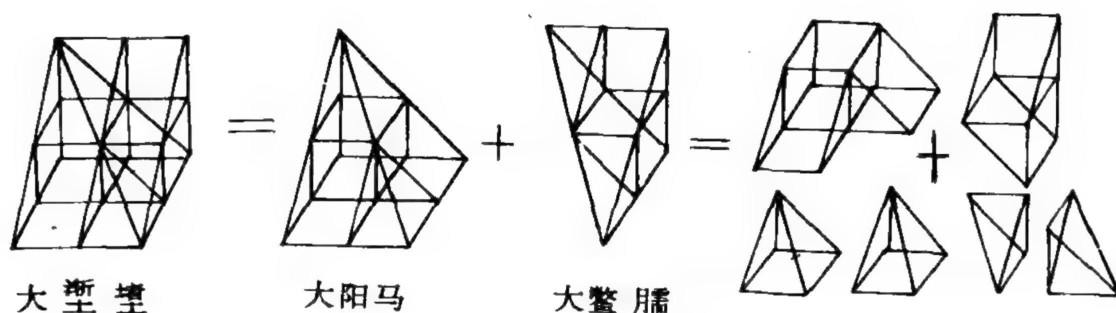


图 3

从图中可以看出

一个大阳马 - 两个大鳖臑 = 2（一个小阳马 - 两个小鳖

臑)。

继续分割，则右边越来越小，最终可以忽略不计。正如刘徽所说：

“半之弥少，其余弥细。至细曰微，微则无形。由是言之，安取余哉。”

对阳马术的更详细的论述可参看 Wagner 的一篇出色论文，发表于 Historia Math. 6 (1979), P169—188。

刘徽还研究了曲面体求积问题，其中有著名的球体积问题。他得出的结论是：球的体积依赖于一个奇特的立体的体积，这个立体是一个立方体的两个内切圆柱的公共部分。刘徽不仅思维精密而且态度严肃，他自己虽然未能解决这个问题，但把这个问题留给后人。他说：“欲陋形指意，惧失正理。敢不阙疑，以俟能言者。”

公元5世纪，大数学家、天文学家和工程学家祖冲之的儿子祖暅，继续了刘徽的工作，终于得到了最后结果。祖暅发现了一条原理，等价于后来重新发现的 Caralieri 原理，即

“缘幂势既同，则积不容异。”

在此不准备用较多篇幅叙述祖暅球体积公式的精采证明。刘徽虽然没有明确提出这一原理，但他在多种简单曲面体体积的推导中确实使用了这一原理。有鉴于此，笔者倡议以“刘祖原理”取代我们中国学者常用的“祖暅原理”的命名。

总而言之，古代中国人以令人陶醉的方式运用出入相补原理、刘徽原理和刘祖原理，这些原理已足以建立起立体——无论是否曲面体——的整个体积理论。

四、代　　数

代数无疑是中国古代数学中最发达的部门。应该指出，中

国古代的代数实际上是方程解法的同义语。探究方程源流，可以发现方程问题是由两个不同方面发展而来的。其一是远古时期的物物交换导致的盈不足术发展而成方程术。《九章算术》的方程章叙述了联立线性方程组的解法，这时引入了负数概念。刘徽在注文中说："并引为行，故谓之方程。"因而可以这样解释："方程"的"方"即是正方形或矩形，"程"是将数字以矩阵形式排列在算板上。

《九章算术》所给出的解联立线性方程组的解法，是进行行列之间的运算，类似于今天的消去法。在刘徽注中还可以看到将数字矩阵变为标准形的详细过程。

方程的第二个来源是测量和几何问题。在求太阳高度时，产生了日高公式和太阳距观测者水平距离的公式。太阳与观测者的距离是利用勾股定理和开平方运算得出的。勾股定理和开平方法的证明都可用出入相补原理推出。《九章算术》第九章中有一问由出入相补原理自然导出了一个二次方程，并将解法命名为"带从开平方法"，这一名称不仅揭示了解此类方程的方法，而且也显露了其根源。

唐朝初年，我国已有了三次方程解法，宋代高次方程的数值解法发展到了高峰，这时的解法与 1819 年霍纳发明的方法本质上完全相同。

宋元时期（10 至 14 世纪）最重要的数学成就是天元概念的引进。天元相当于今天的未知数。尽管在数千年的数学历史发展中，方程解法始终占据着中心地位，但是直到宋元时期才第一次引进未知数的清晰概念和使用系统的方程解法。当时的数学家已认识到了天元术的巨大作用。朱世杰在其著作中写道："今以天元演之，明源活法，省功数倍。"

这一时期，天元术不断发展，最终产生了四个未知数的高次联立方程组的解法，几何代数化、多项式运算和消去法也随

之得到了发展。方程发展的两条主线到此时合二为一，解方程的理论更接近现代代数的有关内容。由于当时所有的运算要在算板上进行，多项式不同类项的系数要摆在板上的不同位置，这就将未知数的个数限制为最多是4。在当时与外部世界如阿拉伯的交流日益加强的情况下，如果放弃算板而采用其他系统，数学将面临一个空前繁荣的时代。但事与愿违，由于种种原因，到元朝末年，深入的研究工作停止了，数学处于停滞不前的境地。当明末利玛窦到中国时，几乎没有一个中国高级知识分子懂得《九章算术》!

五、结 束 语

本文因篇幅所限，不讨论极限概念、高阶内差公式和级数求和等成就。简而言之，中国古代数学的大多数成就具有构造性、算法化和机械化的性质，因此大多数的“术”可以无困难地转化为程序用计算机来实现。中国古代数学家善于从简明的事实得出深刻的结论，并总结成简洁的原理。正是这些简单易明、而应用广泛的原理，形成了中国古代数学的独特风格。中国古代数学的研究方法是从研究具体问题入手，从中提出简单明了的原理和一般方法。代数化和十进位位值制这样杰出的成就也是在这种思想的指导下取得的。总之，中国古代数学有其自身的长处，当然也有传统的不足。我们既不能象明先祖们那样妄自菲薄，漠视我们祖先所取得的成就，亦不可仿效唐人，妄自尊大，拒绝接受外国的先进技术。只要我们充分认识我们传统思想方法的威力，同时吸收外国的先进科学技术，可以预见中国数学将进入一个辉煌灿烂的新时代。

从《数书九章》看中国传统数学构造性与机械化的特色*

在已译成中文斯特洛伊克著的《数学简史》[1]中，有一段耐人寻味的议论：

> 在一切古代东方数学中没有任何地方足以使我们发现我们所谓证明的任何企图。从未用过推理，而仅仅是列出某些规则来："如此做，做这个"。……对于我们这些被欧几里得的严格推理所教育的人，这整个的东方思考方法在最初似乎是惊异而又高度地令人不满。但是当我们认识到我们讲授给我们今天的工程师们和技术人员们的数学大部仍是"如此做，做这个"的方式，而很少有严格的证明企图时，这种惊异就会消灭了。在许多中学中，代数学现仍被教成一堆公式而不是一种演绎的科学。东方数学似乎从未由它所由而生的几千年来技术学和行政问题的影响下解放出来。

斯特洛伊克看出了以中国包括印度在内为代表的所谓东方数学，与以欧几里得为代表的数学有着本质上的差别，而且直到现在东方数学在中学教育中还占据着统治地位，可谓见解精辟独到，值得称道。

但是，对于所谓东方数学缺少推理证明之说，则不能不予以澄清。

---

* 本文摘自《秦九韶与数书九章》(吴文俊主编)，北京师范大学出版社，1987年。

欧几里得创造了一套用定义、公理、定理构成的逻辑推理演绎体系，自有其独到之处，但并不是必须按照这一格式才算是逻辑推理。固然牛顿的《原理》是按照欧几里得的格式来写作的，但笛卡儿的《方法》[3]就没有照此办理，总不能说笛卡儿就没有推理没有证明了。不仅是17、18世纪的大量数学著作没有遵守欧几里得的格式，即使是现代流行的数学刊物，也可以看到不少重要的论文写作是没有遵守这一格式的。

至于我国的数学经典著作，则更是另有一种独特的表达方式，与欧几里得的方式尽管大相径庭，却同样能起到一定的演绎推理与证明的作用，国内论者已多，不再重述。特别值得一提的是李继闵同志的看法，我国古代数学往往寓理于算，不证自明，下文当再说明此点。

其次，公理、定义、定理的表达方式，并不能保证逻辑推理就能严密无间。欧几里得《几何原本》（以下将简称《原本》，参阅〔2〕）在逻辑上弊病甚多，这在19世纪数学批判性浪潮中已多所指摘，即使Hilbert的《几何基础》一书被认为已奠定了所谓欧几里得几何的严密基础，作者近年也已为文指出，欧几里得的证明方式是不可能达到应有的严密程度的，而且是几乎无可挽救的。由于离题过远，在此不再深论。

另一面，我国的古代数学并没有采用欧几里得的演绎形式。但同样达到正确结论，其成果之辉煌，远非同时期世界其他地区的数学可以比拟。象刘徽《九章注》中所表现的推理论证之严密细致，当代数学家如Wagner等也已为之心折。不能否认欧几里得的演绎体系与表达形式有某种优越之处，但不能与推理的严密性视为等同。而我国古算的表达形式有其另外优越的一面，非欧几里得体例所能及，作者以后有机会当专文评述，在此不再深论。

总之，形式应为内容、目的服务，数学的内容实质才是主

要的一面。就内容实质而论，所谓东方数学的中国古代数学，具有两大特色，一是它的构造性，二是它的机械化。

19、20世纪之交，数学应否构造性地发展曾经引起激烈的争论。并因此而深刻影响了当代数学的发展。对于构造性概念的本身，数理逻辑学家还分成了五个不同的层次。对这些论争将由专门家来讨论，非本文所能及。就本文所拟讨论的《数书九章》来说，不妨把构造性与机械化的数学看作是可以直接施用之于现代计算机的数学。我国古代数学，总的说来就是这样一种数学，构造性与机械化，是其两大特色，算筹算盘，即是当时施用没有存储设备的简易计算机。

就《数书九章》而论，我们将就三个方面来举例说明。

一、质数与等数

欧几里得《几何原本》共分13篇，其中7、8、9三篇是数论。质数概念是整个数论的基石，该书第9篇命题20证明了质数的数目比任何指定数目都要多。欧几里得一书原来的证法只就祗有三个质数立论，叙述也比较繁琐，现在已成经典的证法，则如下述。假定只有有限个质数 p_1，p_2，…，p_n，作出 $p_1p_2\cdots p_n+1$，于是不论这数本身是否质数，都将引出矛盾，因而有任意大的质数，如所欲证。

这是一种典型的非构造性的证法，只证明了任意大质数的存在，但对于如何给出比一个数更大的质数来并未提供任何线索。

在我国，则根本没有质数的概念，自然谈不上奠基在质数概念上的数论等分支学科的形成。

但是，数学决不能认为是单纯作为推理训练之用的学问，更不能沦为数字游戏之学。数学之所以重要，主要是因为能有

助于人类认识自然、控制自然，为解决应用中产生的各种问题提供有效手段。就这一点来说，我国古代没有质数的数论要比欧几里得的数论不仅毫无逊色，而且要优越得多。《数书九章》中的大衍求一术，就是我国古代数论的一项代表性的成就。

大衍求一术，向被誉为中国剩余定理，但在国外文献阐述中国剩余定理时，往往局限于诸定母两两互质的特殊情形，而在原来的大衍求一术中，却没有这一限制。在大衍求一术中，并不需要质数与互质这些概念，而代之以连环求等的方法，巧妙地得到了一般情形的解法。

求等之等的等数，即是现代通称的最大公约数。在《原本》第7篇的第1，2，3三个命题，即是关于求若干整数（事实上不超过3个）的最大公约数，这在后世被称为欧几里得算法。

在我国，求两数最大公约数即等数用更相减损之术，将两数以小减大累减以得之。如求24与15的等数，其逐步减损如下表所示：

$$(24,15)\to(9,15)\to(9,6)\to(3,6)\to(3,3)$$

每次所得两数与前两数有相同的等数，而两数之值逐步减少，因而到有限步后必然获得相同的两数，也即所求的等数，其理由不证自明。(参阅沈康身同志的有关著作〔4〕。)

这个寓理于算不证自明的方法，是完全构造性与机械化的，尽可以据此编成程序上机实施。

欧几里得算法与上述更相减损之术实质上是相同的，但《原本》除了引进许多概念、定义以外，还给出了一个证明。这个证明占用了足足三页的篇幅（Heath的注释已除去不算），不仅繁长艰涩，而且依据Heath有下面的注释：

> 请注意欧几里得假定了下面的事实作为一个公理，即如果a，b都被c除尽，则$a-pb$也能被c除尽。在下一

命题中，他也假定了一个公理，即在同样情形下 c 也能除尽 $a+pb$。

这个注释无异是说，欧几里得的原证在逻辑上并不严密而且有漏洞，它隐含了一些不该略去的公理，岂仅是证明繁长晦涩而已。

《原本》第 7 篇的命题 1 与 2，虽然等价于后世通称的欧几里得算法，但就其证明原来的叙述方式来说，既不是构造性的，也不是机械化的。即使按照后世已经转化加工成真正算法的形式，在整个欧几里得体系中，也只能算是一个例外。这与贯穿在我国古代数学中构造性与机械化的算法成果相比，是不可同日而语的。

我国古代数学中有等数而没有质数，没有质数该怎么办？莫绍揆同志已有一文[5]论之，甚为精辟，定母完全任意的大衍求一术提供了不用质数概念的又一实例。

在任意定母的情形下秦九韶应用连环求等的术文虽少有瑕疵，但不难纠正，稍加改动即可，可以参阅本书有关论文。

明代以来我国数学日益衰微，九章之学几已湮没无闻。一自利玛窦东来，数学舞台遂为欧几里得之学所垄断独占。清代虽不乏有志有为的杰出之士，致力于发掘古籍发扬古学，但终究因衰微日久，往往只知以欧几里得之学生搬硬套，对古籍强为穿凿作解。有计算经验者都知道要把一数分解成质数因子并非易事，而求两数的等数却要容易得多。清张敦仁等以质数概念解释定母任意的大衍求一术，既使原意全失，又无异于化易为难，我认为不足为法，是不可取的。

二、增乘开方法与正负开方术

从《九章》中开平、立方发展至宋元时期增乘开方法与正

负开方术的求方程数值解法，是中国古代数学构造性与机械化思想方法的又一代表性成就。

为简单起见，试考虑下面形式的方程

$$x^n + a_1 x^{n-1} + \cdots + a_{n-1} x = a_n, \tag{1}$$

其中 a_i 都 >0。计算数学应讲求正确与效率，我们试就这种方程对宋代古法作一检查。检查方法将使用一个简单的计算器对这种方程求解来作出判断。

作者曾用一种 HP25 型的袖珍计算器做过这种检查工作。这种 HP25 有 8 个存储单元可以存放数值，利用这 8 个单元就可以编一个小程序解高至 5 次的方程（1），而且可达到任意预定的精度，所用方法在于利用我国独创的十进位位值制数字表达法逐步求得解的各位数字。例如对 $x^2=2$ 求 2 的平方根时，将逐步求出 1，1.4，1.41，1.414，1.4141，…其每次修正数值为 0.4，0.01，0.004，0.0001，而最后到 1.4141 为止时所达精度为 0.0001。

今以 3 次方程（1）为例，利用 HP25 编程序如下。

HP25 的 8 个存储单元记作$\boxed{0}$，$\boxed{1}$，…，$\boxed{7}$。我们将所设方程系数 a_1，a_2，a_3 各存入$\boxed{1}$，$\boxed{2}$，$\boxed{3}$，逐次求得的方程近似解设为 x_0 存入$\boxed{0}$，逐次解的修正数值记为 c 存入$\boxed{6}$，即每次近似解 x_0 改进为下一次的 x_0+c，又所预定要求的精度设为 $\delta=0.0_k1$ 存入$\boxed{7}$。这里没有用到存储单元$\boxed{4}$，$\boxed{5}$，说明本例还没有使用上 HP25 的全部潜力，实际上同样可以解 5 次方程。

先将 a_i，c，δ，以及尝试解 x_0 存入$\boxed{1}$，$\boxed{2}$，$\boxed{3}$，$\boxed{6}$，$\boxed{7}$，$\boxed{0}$后，即可依照下面的程序逐步按键求得所要求精度的数值解，程序右面部分是简单的说明。

步骤	程　　序	说　　　明
1	*RCL*0	将$\boxed{0}$中的 x_0 取出 3 次
2	*ENTER* ↑	
3	*ENTER* ↑	
4	*ENTER* ↑	
5	*RCL*1	依 $x_0^3+a_1x_0^2+a_2x_0-a_3=((x_0+a)x_0+a_2)x_0-a_3$ 的计算步骤将所得存显示中
6	+	
7	×	
8	*RCL*2	
9	+	
10	×	
11	*RCL*3	
12	−	
13	$g:x=0$	显示中数＝0 时
14	*GTO* 32	
15	$g:x<0$	显示中数＜0 时
16	*GTO* 26	
17	1	显示中数＞0 时 将$\boxed{6}$中 c 改为$c/10$
18	0	
19	*STO*÷δ	
20	*RCL*6	比较$\boxed{6}$中的 c 与$\boxed{7}$中的 δ，相减所得 $c-\delta$ 存显示中
21	*RCL*7	
22	−	
23	$g:x\geqslant 0$	显示中数≥0 即 $c\geqslant\delta$ 时
24	*GTO* 28	
25	*GTO* 34	显示中数＜0 即 $c<\delta$ 时
26	*R*↓	将$\boxed{0}$中 x_0 修改为 x_0+c 并返回第 2 步重新开始
27	*STO*0	
28	*RCL*0	
29	*RCL*6	
30	+	
31	*GTO* 2	
32	*R*↓	$\boxed{0}$中的数即所求解并在显示中展出
33	*STO*0	
34	*RCL*0	

用这一方法求方程实解所依据的原理，从计算过程即已完全清楚，已可不加辞费不证自明，而且所得结果是绝对精确的（指到预定精度而言），用不着顾虑是否收敛以及收敛快慢的问题。

对于一般形式的方程，可用益积减从正负开方等术，其理相同，可参见本书有关论著。

我国古算只涉及求实根，其实求复根也一样可以。因为每一个复系数求复根的方程，相当于两实系数的实变方程组。若变量为 x、y，可依宋元时期消元的方法得到一只有 x 的实系数方程，即可依增乘开方与正负开方之术求得 x、y 的实数解，也就是原来复方程的复数解。

目前通行的方程求数值解法是 Newton 法一类的迭代法，这种方法在初始值的选取，收敛与精度，系数稍有偏差时的影响，都存在着复杂甚至颇为严重的问题。当前 Berkeley 的 Smale 与 Princeton 的 Kuhn 都从不同的角度对 Newton 法进行了理论探讨，可见其问题之不简单。与我国古法相较，古法之直捷了当且可排除各种病态的优越性是颇为显著的。

三、大衍求一术

大衍求一术是我国古代数学构造性与机械化思想方法的又一范例。这在国外称为中国剩余定理，但往往只涉及定母两两互质的情形，用现代语言可表述如下：

设有 n 个两两互质的正整数 m_1，…，m_n，又设 u_1，…，u_n 为已知整数，则有唯一的一个 $\geqslant 0$ 而 $< m = m_1 m_2 \cdots m_n$ 的整数 u 使

$$u \equiv u_j \bmod m_j, \qquad 1 \leqslant j \leqslant n。$$

在 Knuth[6] 一书中，对中国剩余定理给出了两个证明，一个是非构造性的，另一个是构造性的。这两证明颇可说明构造性与非构造性的区别，故录之于此。

非构造性证明。唯一性的证明是容易的，故只须证明存在性即足。为此，对任一 $n>0$ 记 $\geqslant 0$ 而 $<n$ 的整数全体为 $[n]$，又对任一整数 v，记 $[m_j]$ 中与 v 对 m_j 同余的唯一整数为 v_j，考虑对应关系

$$\lambda:\begin{cases}[m]\longrightarrow[m_1]\times\cdots\times[m_n],\\ v\longrightarrow(v_1\cdots,v_n)。\end{cases}$$

显然 $[m]$ 中不同的 v 将对应不同的 $(v_1,\cdots,v_n)$，因集合 $[m]$ 与 $[m_1]\times\cdots\times[m_n]$ 中的个数相同，故 λ 是一对一的。特别对于每一组 $[m_1]\times\cdots\times[m_n]$ 中的数，例如 $(u_1,\cdots,u_n)$ 必有一 u 与之对应，这就是要证明的定理。

与前面所引《原本》中质数个数无限这一定理的证明相仿，这是一种典型的非构造性证明。这种证法简单美妙，但对要求得的 u 并不提供任何求法的线索。构造性的证明则不然，不仅证明 u 的存在，同时也给出了 u 的求法甚至显豁公式。下面的证明就属于这一种，也在 Knuth 同书中给出。

构造性证明。对任意数 n，命 $\varphi(n)$ 是它的 Euler 数。作

$$M_j=(m/m_j)^{\varphi(m_j)},\qquad 1\leqslant j\leqslant n,$$

则据整数论定理，任意满足

$$u\equiv(u_1M_1+\cdots+u_nM_n)\bmod m$$

的数 u 都能使 $u\equiv u_j \bmod m_j$，$1\leqslant j\leqslant n$，于是只须取 $\geqslant 0$ 而 $<m$ 的 u 即合所求。

证明中的 $\varphi(n)$ 是 Euler 函数，即 $\leqslant n$ 而与 n 互质的正整数个数。如果 n 分解成质因子的乘积

$$n = p_1^{e_1} \cdots p_s^{e_s},$$

其中 p_j 是不同质数，则 Euler 证明

$$\varphi(n) = n\left(1 - \frac{1}{p_1}\right) \cdots \left(1 - \frac{1}{p_s}\right)。$$

在 n 本身是质数 p 时，

$$\varphi(p) = p - 1, \text{例如 } \varphi(83) = 82。$$

但对一般的 n，则须先分解成质因子再求出 $\varphi(n)$，例如

$$n = 225600 = 2^6 \cdot 3 \cdot 5^2 \cdot 47,$$

故

$$\varphi(225600) = 2^6 \cdot 3 \cdot 5^2 \cdot 47 \cdot \frac{1}{2} \cdot \frac{2}{3} \cdot \frac{4}{5} \cdot \frac{46}{47} = 2^9 \cdot 5 \cdot 23$$
$$= 58880。$$

由于分解因子并非易事，而所得 $\varphi(n)$ 在 n 稍大时即相当大，M_j 因而 u 依上证的计算量将很可观，甚至可使一台现代的计算机也难以处理。

我国古代数学没有质数概念，自然更没有 Euler 函数以及整数理论，但同样解决了这一问题，而且原理浅显，计算简易，甚至解决得更彻底——包括了诸 m_j 不必两两互质的情形。

为简单起见，仍只考虑诸 m_j 两两互质的情形。不论是 Knuth 所载的构造性证法（以下简称西法），还是秦九韶的大衍求一术，其基本原理可以说都如下述：

假定有一组整数 N_j，使

$$N_j \equiv \begin{cases} 1 \bmod m_j, \\ 0 \bmod m_i, \; i \neq j \end{cases} \qquad (1 \leqslant j \leqslant n),$$

则任意使

$$u \equiv (u_1 N_1 + \cdots + u_n N_n) \bmod m$$

的整数 u 都能满足同余式

$$u \equiv u_j \bmod m_j, 1 \leqslant j \leqslant n。$$

在西法中，所取 N_j 是

$$N_j = M_j = (m / m_j)^{\varphi(m_j)},$$

其依据是整数论中的一条定理，其中 $\varphi(m_j)$ 的概念与公式以及定理的证明都是颇费周折的。

在秦法大衍求一术中，则是先定 k_j，使

$$k_j(m / m_j) \equiv 1 \bmod m_j,$$

再取 $$N_j = k_j(m / m_j),$$

计算简单而理论依据已寓于计算之中，可以不证自明。

在大衍求一术中，所用专门术语是

m_j——定母，$m = m_1 m_2 \cdots m_n$——衍母，

m / m_j——衍数，(m / m_j) 对 m_j 之余 n_j——奇数，

k_j——乘率，$k_j(m / m_j)$——用数，

u_j——余数，u——总数。

这些数之间的关系是

$k_j * n_j \equiv 1 \bmod m_j$，或乘率 * 奇数≡1mod 定母，

$u = \sum u_j * k_j(m / m_j)$，或总数 = ∑余数 * 用数。

求一术的最主要一步在于从奇数与定母求乘率，即已知 n_j，m_j 时求 k_j，即解

$$k_j * n_j \equiv 1 \bmod m_j。$$

解法的出发点是考虑更一般形式的同余式

$$x * b \equiv c \bmod a,$$

其中 a，b，c 都是已知整数而 x 为所求，在 $c = 1$ 而 b，a 各为 n_j，m_j，所得 x 即为乘率 k_j。

上面的同余式有两个明显的解答：

$c = b$ 时，$x = 1$，即 1 * 奇数≡奇数，

$c = -a$ 时，$x = 0$，即 0 * 奇数≡一定数，

而在　　$c=1$ 时，x 应 = 乘率。

上面两个显然的同余式可以把其中出现的四数置于一个方阵的四角来简单地表示，即置：

$$\begin{array}{|cc|}\hline 左上 & 右上 \\ 左下 & 右下 \\ \hline\end{array} = \begin{array}{|cc|}\hline 1 & 奇数 \\ 0 & 定数 \\ \hline\end{array}$$

方阵四角的数符合下面的关系式：

左上数 * 奇数≡右上数 mod（定母），

左下数 * 奇数≡ - 右下数 mod（定母）。

开始时位于方阵右方上下的奇定二数可以很大，求一术将四角之数递互累乘累减，使右方两数逐步减小，但却依然保持上面两个关系式成立。到右上角之数减小至 1 时，左上数由于上关系式就给出了所求的乘率。

将四角之数递互累相乘减以减小右方二数而保持两关系式之法，实质上与古法求两数的等数时以少减多更相减损以减少两数之值，而使等数重叠不变的更相减损之术理由是相同的，只是更为复杂而已。这可以说是求一术的真髓。

现取《数书九章》中余米推数题为例来对求一术作一具体说明。该题是一件盗窃案的侦破术，原题照录如下。

> 问有米铺，诉被盗去米一般三箩，皆适满，不记细数。今左壁箩剩一合，中间箩剩一升四合，右壁箩剩一合，后获贼，系甲乙丙三名。甲称当夜摸得马杓，在左壁箩，满舀入布袋，乙称踢着木履，在中箩，舀入袋，丙称摸得漆椀，在右边箩，舀入袋，将归食用，日久不知数。索到三器，马杓满容一升九合，木履容一升七合，漆椀容一升二合。欲知所失米数，计赃结断三盗各几何。

今以合为单位，则依题意

剩米数＝余数 $u_j=1$， 14， 1，

盗器容量＝定母 $m_j=19$， 17， 12，

衍母 $m=m_1m_2m_3$， $19*17*12=3876$，

衍数 $m/m_j=204$， 228， 323，

奇数＝m_j 除 m/m_j 之余＝14， 7， 11，

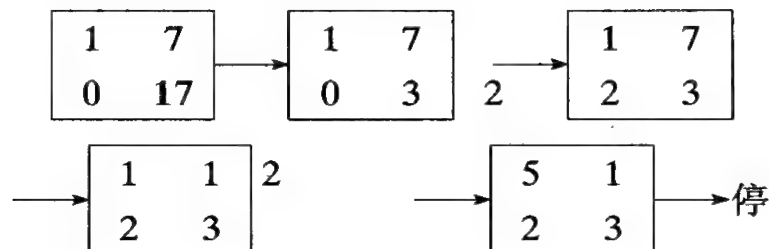
所求乘率＝$k_j=15$， 5， 11。

这里乘率 k_j 的求法依大衍求一术如下表所示：

步骤	大衍求一术云	
1	置奇右上、定居右下，立天元一于左上。	1 14 / 0 19
2	先以右上除右下，所得商数。	1 14 / 0 5 （商 1）
3	与左上一相生入左下。	1 14 / 1 5
4	然后乃以右行上下、以少除多，递互除之，所得商数。	1 4 / 1 5 （商 2）
5	随即递互累乘，归左行上下。	3 4 / 1 5
6	（同第 4 步）	3 4 / 1 1 （商 1）
7	（同第 5 步）	3 4 / 4 1
8	（同第 4 步）	3 1 / 4 1 （商 3）
9	（同第 5 步）	15 1 / 4 1
10	须使右上末后奇一而止	停
11	乃验左上所得以为乘率	乘率 $k_1=15$

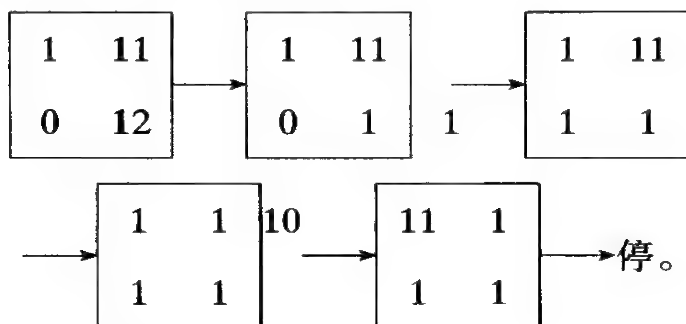
其余两个乘率的求法也一样，算式如下：

求 k_2，使 $k_2 * 7 \equiv 1 \bmod 17$，



得 k_2 = 左上数 5。

求 k_3 使 $K_3 * 11 \equiv 1 \bmod 12$。



得 k_3 = 左上数 11。

计算乘率 k 的过程颇为简单，且其原理正与求等数的原理类似，颇为显然，即所谓理在算中，不用多加辞费，已可不证自明。

若用西法，其计算量就要大得多。试将两种方法列表比较如下。

求

$$u \equiv \begin{cases} 1 \bmod 19, \\ 14 \bmod 17, \\ 1 \bmod 12。 \end{cases}$$

在《数书九章》所列大衍求一术的 9 个问题中，余米推数中的数都不大，但西法的计算量已颇可观。在另一题古历会积中，出现了

奇数 = 9253，定母 = 225600，

西　　法	秦法（大衍求一术）
$N_j=(m_1\cdots m_n/m_j)^{\varphi(m_j)}$	$k_j*(m_1\cdots m_n/m_j)\equiv 1\bmod m_j$ $N_j=k_j(m_1\cdots m_n/m_j)$
$N_1=(17*12)^{\varphi(19)}=204^{18}$	$k_1*(17*12)\equiv k_1*14\equiv 1\bmod 19$ $\Rightarrow k_1=15$ $N_1=15*(17*12)$
$N_2=(19*12)^{\varphi(17)}=228^{17}$	$k_2*(19*12)\equiv k_2*7\equiv 1\bmod 17$ $\Rightarrow k_2=5$ $N_2=5*(19*12)$
$N_3=(19*7)^{\varphi(12)}=323^4$	$k_3*(19*17)\equiv k_3*11\bmod 12$ $\Rightarrow k_3=11$ $N_3=11*(19*17)$
$u\equiv 204^{18}+14*228^{17}+323^4\bmod 3876$	$u\equiv 15*(17*12)+5*(19*12)+11*(19*17)\bmod 3876$

按大衍求一术计算，易从

$$k*9253\equiv 1\bmod 225600$$

算得

$$k=172717。$$

若依西法，则须计算

$$N=9253^{\varphi(225600)}=9253^{5880},$$

即使一台现代化的计算机，也是难于完成这一计算任务的。

最后我们想指出，载于 Knuth 一书的西法，自然早已有之，但是由于用到 Euler 函数，因而此法出现的最早年限是 18 世纪。至于大衍求一术，则即使算完全是秦九韶一人的创造，其出现的最低年限，也至迟是 13 世纪——1247 年。

参 考 文 献

〔1〕D. J. 斯特洛伊克，数学简史（关娴译），科学出版社，(1956)。

〔2〕 T. L. Heath, The thirteen books of Euclid's Elements. Dover (1956).

〔3〕 R. Descartes. La géomètrie, Dover (1954).

〔4〕 沈康身，更相减损术源流，全国数学史讨论会，(1981)。

〔5〕 莫绍揆，假如没有素数概念该怎么办？数学研究与评论，4 (1982) 183。

〔6〕 D. E. Knuth, The art of computer programming, Vol. 2, Seminumerical algorithms (1969).

我国古代测望之学重差理论评介兼评数学史研究中某些方法问题*

一、引　　言

《中国古代数学对世界文化的伟大贡献》一文曾经指出：

"代数学无可争辩地是中国的创造，……甚至可以说在16世纪以前，除了阿拉伯某些著作之外，代数学基本上是中国一手包办了的"。

"中国的劳动人民……实质上达到了整个实数系统的完成。特别是自古就有了完美的10进位值制的记数法。……。这一创造对世界文化贡献之大，如果不能与火的发明相比，也是可以与火药、指南针、印刷术一类发明相媲美的"。

不仅在数学系统的完成上或是在代数学的创立上，就是在几何学上，我国古代也有着极其辉煌的成就。我国古代的几何学，立足于广大劳动人民的丰富实践经验，从天文观察与工农业建设中发现问题，提出问题，抓住了几何学的核心与实质，建立了具有我国特色的几何学体系。既有丰硕的成果，又有系统的理论，其内容有许多是微积分得以创立的关键所在，是希腊的几何学所未能做到的。但是大多数西方数学史家对几何学一向标榜欧几里得体系，推崇备至，而对中国古代的几何学则以为不足道，甚至认为一无所有。对此，我们有着截然不同的

---

* 本文摘自《科技史文集》（第8辑），上海科学技术出版社，1982年。

看法。我国的数学工作者们，也应该认真调查、弄清事实以明辩是非，而不能把西方言论视为当然。

恩格斯曾经说过：

“和其它科学一样，数学是从人的需要中产生的；是从丈量土地和测量容积，从计算时间和制造器皿产生的。”

我国古代人民，正如其他各地的许多古代民族一样，观天测地，从事土地的丈量、容积的测量、时间的计算等等生产活动，在此基础上创造了我国固有的古代几何学。从远古时期起，即有了一般形式的勾股定理，并应用之于测日之高远大小，具见《周髀算经》一书。秦汉时期又发展了勾股理论并导致二次方程，成为我国一千多年代数学蓬勃发展的主要源泉之一。到魏晋时期，《周髀算经》以之观天者，刘徽以之测地，建立了测高望远的重差理论。另一面，土地的丈量与容积的测量产生了面积与体积理论，并提炼出出入相补这一一般原理，到5世纪南北朝时期又提出了祖暅原理。这些一般原理的建立，说明了我国古代人民不仅能紧密联系实际，善于分析问题解决问题，而且有着高度的抽象与概括能力。不仅脱离实际把量度排斥于几何学之外的欧几里得所无法比拟，甚至也是虽重视实际但偏重技术缺少概括的阿基米德与海伦所不及。这些原理，不仅使欧洲直至19世纪末才弄清楚的体积问题早在祖暅时代即已具备了完满的基础，且为后世微积分的建立创造了条件，而且，与希腊欧几里得几何的形数割裂者恰恰相反，我国在数学发展过程中自始至终是把空间形式与数量关系融合在一起的，因而数系统的建立与臻于完备，以及代数学的发生发展，也始终与几何学的发展贯穿在一起。到宋元之世天元——也即未知数概念的明确引入，代数式与其代数运算的阐明，以及几何代数化方法的逐渐成熟，更为解析几何的创立开辟了道路。诚如上篇文中所说：

"……到宋元之世，已经具备了西欧17世纪发明微积分前夕的许多条件。不妨说我们已经接近了微积分的大门"。

由于我国数学上成就之极端重要与丰富多采，即使只局限于几何学范围，也决非少数几篇文章就能阐明。本文将只就几何学成就的一个角落——重差理论作一分析介绍。抛砖引玉，希望其他同志能从各个不同方面与角度进一步深入阐发。

二、重差理论历史

由于中国古代经典著作中史料不为一般读者所知，流传不广又不易找到，故把有关重差理论的片段罗列于下，作为全文讨论的依据。

重差理论来源于量日之高远，而量日高远法主要见于公元前50至100年间成书的《周髀算经》，有关段落（《淮南子》（公元前120年左右）与《周礼·地官》也有类似的记述）如下：

"周髀长八尺，夏至之日晷一尺六寸。髀者，股也，正晷者，句也。正南千里，句一尺五寸，正北千里，句一尺七寸。……候句六尺。……故以句为首，以髀为股，从髀至日下六万里而髀无景。从此以上至日，则八万里。若求邪至日者，以日下为句，日高为股。句、股各自乘，并而开方除之，得邪至日。从髀所褰至日所十万里。"

除文字记载外，至今河南登封县告成镇即古阳城，还有周代观景台的遗址〔丙3〕。

三国吴人赵爽为《周髀算经》作注，作有《日高图》与《日高图说》。《日高图说》全文如下：

"黄甲与黄乙其实正等。以表高乘两表相去为黄甲之实。以影差为黄乙之广而一，所得则变为黄乙之袤，上与

日齐。按图当加表高。今言八万里者，从表以上复加之。青丙与青己其实亦等。黄甲与青丙相连，黄乙与青己相连，其实亦等。”

赵爽日高图的原图早已遗失。现传各本中原图俱误，如图1。钱宝琮依据该图与赵爽注文重绘如图2。

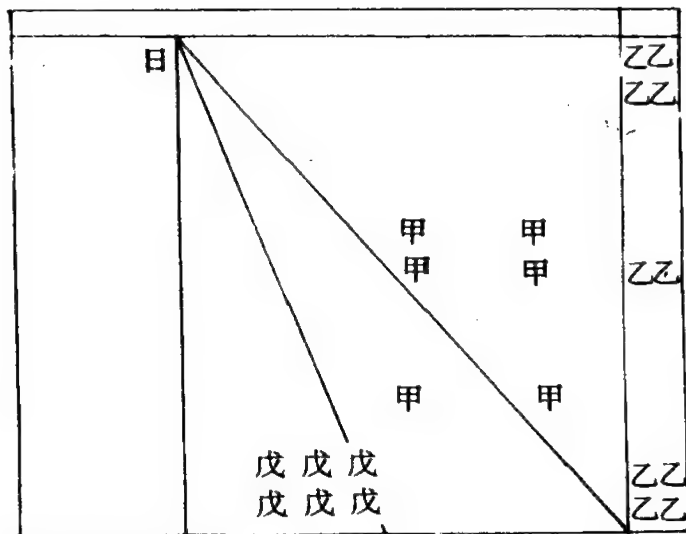


图1

公元3世纪魏刘徽作《九章算术注》。序中说：

“……〈周官〉大司徒职，夏至日中立八尺之表，其景尺有五寸，谓之地中。说云，南戴日下万五千里。夫云尔者，以术推之。按《九章算术》立四表望远及因木望山之术，皆端旁互见，无有超邈若斯之类。然则苍等为术犹未足以博群数也。徽寻九数有重差之名，原其指趣乃所以施于此也。凡望极高、测绝深而兼知其远者必用重差。句股则必以重差为率，故曰重差也。立两表于洛阳之城，令高八尺。南北各尽平地，同日度其正中之景。以景差为法，表高乘表间为实，实如法而一，所得加表高，即日去地也。以南表之景乘表间为实，实如法而一，即为从南表至南戴日下也。以南戴日下及日去地为句股。为之求弦，即日去人也。……虽天圆穹之象犹曰可度，又况泰

山之高与江海之广哉。徽以为今之史籍且略举天地之物，考论厥数，载之于志，以阐世术之美。辄造重差，并为注解，以究古人之意，缀于句股之下。度高者重表，测深者累矩，孤离者三望，离而又旁求者四望。触类而长之，则虽幽遐诡伏，靡所不入。博物君子，详而览焉。”

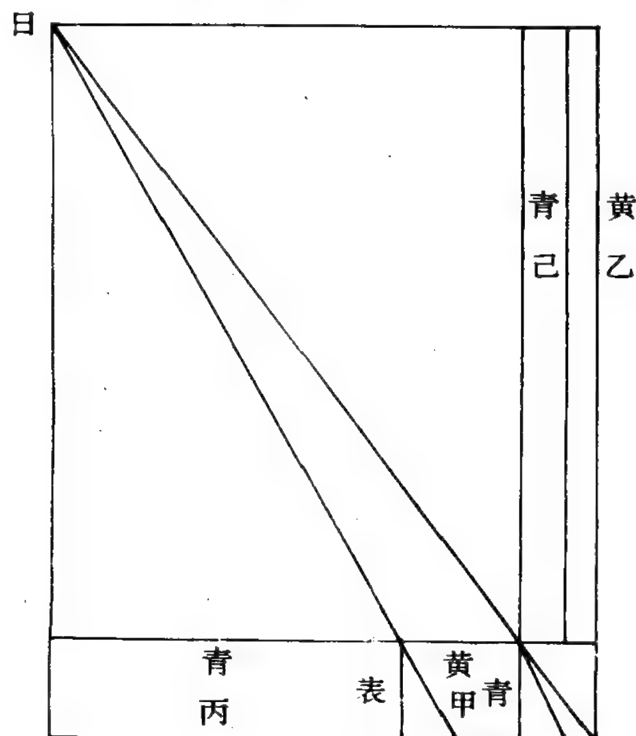


图 2

这一序中概括了日高理论中的三个基本公式（《周髀算经》与刘徽《九章算术注》中求日径的部分都已略去）：

$$日去地=\frac{表高\times表间}{景差}+表高,$$

$$南戴日下=\frac{南表景\times表间}{景差},$$

$$日去人=\sqrt{(南戴日下)^2+(日去地)^2}。$$

刘徽把这一量日高的理论施之于量“泰山之高与江海之广”，“辄造重差，并为注解，以究古人之意，缀于句股之下”。

原来这一重差理论是作为《九章算术注》十卷中的第十卷附于勾股章之后的。唐代把第十卷重差和九章分离而另本单行，改称《海岛算经》。原作有注有图。但此后注图都已失传。现存《海岛算经》只剩九题，其中第一题望海岛照录如下：

“今有望海岛。立两表齐高一丈，前后相去千步。令后表与前表参相直。从前表却行一百二十三步，人目著地取望岛峰，与表末参合。从后表却行一百二十七步，人目著地取望岛峰，亦与表末参合。问岛高及去表各几何。

术曰：以表高乘表间为实。相多为法，除之。所得加表高，即得岛高。

求前表去岛远近者，以前表却行乘表间为实，相多为法，除之，得岛去表里数。”

按我国古算书中的“术”可泛指方法、定理、公式或理论。这里的“术”所涉及的理论可概括为下面两个基本公式：

$$\text{岛高}=\frac{\text{表高}\times\text{表间}}{\text{相多}}+\text{表高},$$

$$\text{前表去岛}=\frac{\text{前表却行}\times\text{表间}}{\text{相多}}。$$

这两公式也即前面关于日高的基本公式，也可与《周髀算经》以及赵爽《日高图说》相对照。

《海岛算经》其余八题依次是：

2．望松　3．望邑　4．望谷　5．望楼　6．望波口　7．望清渊　8．望津　9．临邑。诸题不论是重表、累矩、三望、四望，都是从不同测点获得不同的测望数据，相互较核，取其差值为法而给出明显公式。所依据重测取差的方法原则各题是一致的，与重差之名正相符合。为此，我们将把这一测望理论称为重差理论，所得诸公式统称为重差公式。在各个特殊情形则将称为日高公式、海岛公式、松高公式等等。

三、后代关于海岛公式的证明

刘徽的《海岛算经》原来有注有图。“析理以辞，解体用图”，所谓注即相当于现代的分析与证明。可惜注图都已遗失。后世以迄近代有不少中外人士曾补作证明。现择其重要者依历史先后次序介绍于下，再于下节讨论其正误得失。

现存最早讨论《海岛算经》诸题的证明者见于宋杨辉的《续古摘奇算法》(公元 1275 年)。杨于书末有海岛题解云：

“魏刘徽注九章，立重差，著于句股之下，以阐世术。夫度高测深，句股之法则，无自而可知。故重表、累矩、三望、四望，旁求审察。是以松山高下，方邑大小，其重表也。岸望谷深，山望津广，其累矩也。登望松高，遥望波口，非三望之术乎。清渊白石，登山临邑，非四望之术乎。海岛去表为之，篇首因以名之，实九章句股之遗法也。”

“迄今千余载间，唐李淳风而续算草，未闻解白作法之旨者。辉尝置海岛小图于座右，乃见先贤作法之万一。若欲尽传，岂不轻易秘旨。或不传流，亦无以伸前贤之美。本经题目广远，难于引证，学者非之，今将孙子度影量竿题问，引用详解，以验小图。姑以一问，其余好学君子自能触类而考，何必轻传。”

杨辉的另一现存著作《乘除通变本末》上卷（1274 年）中也说：

“海岛题法，隐奥莫得其秘。李淳风虽注只云下法，亦不曾说其源。《议古根源》（12 世纪刘益著，现已失传）原无细草，但依术演算，亦不知其旨。”

观此，杨辉认为没有必要给出《海岛算经》九问的全部证明，而着重在望海岛第一题的详细分析：“姑以一问”，“解白

作法之旨”，而“说其源”，以“见先贤作法之万一”。其余则留待“好学君子”“触类而考”。从“尝置海岛小图于座右”与“以验小图”等语，可能杨辉所见即系刘徽原来的“海岛小图”，至于现存杨辉的原图则大体如图3。杨的分析是把图3分解成两个小图即图4、5，每图相当于一个较简单的孙子度影量竿问题。杨并指出证明的依据可归结为下述原理，即杨所说的“源”：

“凡股中容横，句中容直，二积皆同。”换言之，即图4、5中的

面积甲＝面积乙。

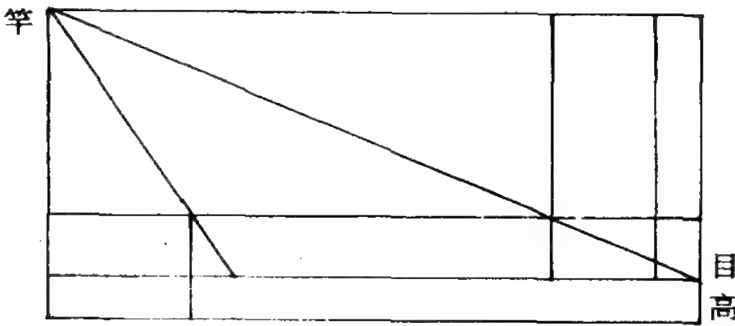
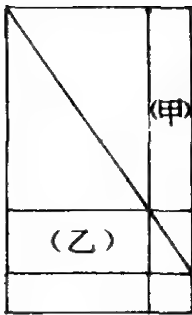


图3



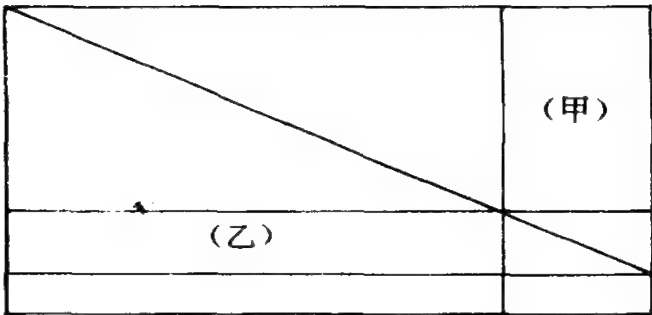


图4　　图5

明末（16、17世纪之交）西方殖民国家的先遣部队耶稣教会教士为便于传教以奴役中国人民，借宣扬数学历法等科学

文化之名而进入明室宫庭。由于明季八股取士，理学统治了学术界的思想，科学技术的发展受到摧残，当时中国古代的传统数学几已失传。士大夫如徐光启等为传入的西方天文历算所惑，与来华教士利玛窦等争相结纳，除翻译《几何原本》外，又译《测量法义》(1607—8) 一书。《测量法义》共 15 题，书前题有利玛窦口授，徐光启笔受。该书主要介绍欧洲的测量方法，而其中第 10 题以表测高的后面部分所谓“欲测甲乙之高，……若戊不欲至乙或不能，则用两表较算……，”其内容即与《海岛算经》第一题无异（图 6)。

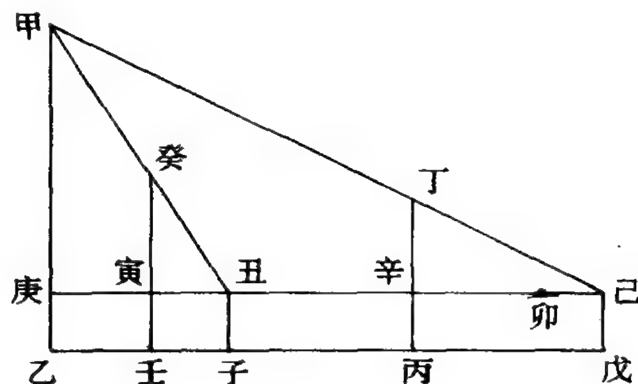


图 6

在《测量法义》之后，徐光启又著《测量异同》。序言中说：

“九章算法句股篇中故有用表用矩尺测量数条。与今译《测量法义》相较，其法略同，其义全阙，学者不能识其所由。既具新论，以考旧义，如指掌矣。今悉存诸法，对题胪列，惟求同异，以竢讨论。其旧篇所有，今译所无者，仍补论一则，共为测量异同六首如左。”

《测量异同》中所列 6 首中第 4 首以重表兼测无远之高、无高之远以及第 6 首以重矩兼测无广之深、无深之广，即《测量法义》中的第 10 题后面部分也即《海岛算经》的第一题。

不论是利玛窦口授的《测量法义》或是徐光启补作的《测量异同》，在证以表测高题即海岛公式时，都用相似三角形与比例，并引述《几何原本》中的命题。《测量法义》中的证明依自戊向前抑退后至子而分成两种情形，又比较两影的长短，叙述颇为繁琐。《测量异同》证明与此类似。现将《测量法义》中证明的主要部分介绍如下，见附图 6 戊向前至子的情形，其中丙丁与壬癸为表，己、丑为人目：

由相似三角形得	己辛:辛丁＝己庚:庚甲，
交换比例项得	己辛:己庚＝辛丁:庚甲。
同样	丑寅:丑庚＝寅癸:庚甲，
因辛丁＝寅癸，故	丑寅:丑庚＝辛丁:庚甲，
在己庚上取点卯，使	己卯＝丑寅，
则得	己辛:己庚＝己卯:丑庚。
由比例定理，得	己辛:己庚＝卯辛:己丑，
与前第二等式比较，得	卯辛:己丑＝辛丁:庚甲，
交换比例项得	卯辛:辛丁＝己丑:甲庚，
或	卯辛:辛丁＝子戊:甲庚，
即	两测较:表目较＝距较:甲庚。

由此得甲庚之高。

加目至足之数，即得甲乙之高。

在清初数学这一阵地实质上已为西算所完全占领，我国古代传统数学几成绝学。直至乾隆三十八年（公元 1773 年）开始编辑《四库全书》，古典数书陆续发现，引起了研究古典数学的高潮。19 世纪初李潢作《九章算术细草图说》与《海岛算经细草图说》，为《九章算术》注疏作图并补出证明。李的《重差图序》说：

“图九。望海岛旧有图解，余八图今所补也。同式形两两相比，所作四率，二、三率相乘与一、四率相乘同积。如欲作

图明之，第取一、三率联为一边，又取二、四率联为一边。作相乘长方图之，自然分为四幂。又以斜弦界为同式句股形各二，则形势验矣。旧图于形外别作同积二方，至两形相去辽远者，又必宛转通之，皆可不必也。图中以四边形五边形立说，似与句股不类。……”

从李潢序可见李也曾见过“望海岛旧图”。可惜现存李潢著作中“旧图”已无，而代之以李所补作之图。现将李所作海岛题的图（图 7）与证明大意概述如下：

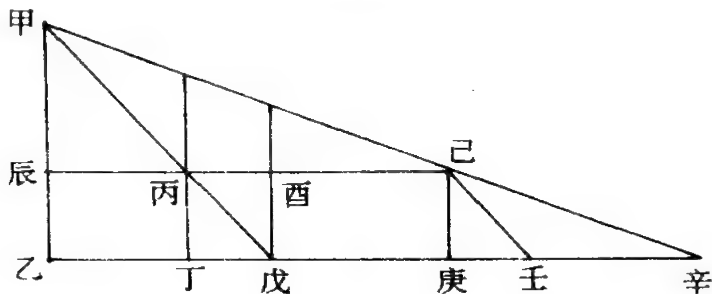


图 7

在图 7 中甲乙为海岛，丙丁与己庚为前后两表，戊与辛为着地人目。作己壬与丙戊平行，则因△己庚辛≈△甲辰己，与△己庚壬≈△甲辰丙，得辛壬：己庚＝己丙：甲辰。故

$$甲辰=\frac{己庚\times 庚丁}{辛庚-戊丁}，即得岛高公式。$$

若与利玛窦、徐光启的证明比较，可见李多添了一条平行线，至于其他诸题的证法也类似，但添线更多，推理更繁。

近几十年对《海岛算经》诸题的证明则往往应用代数。例如三上义夫在《中日数学史》一书介绍《海岛算经》时说（图 8）：

海岛术文相当于公式

$$山高\ x=a+\frac{ab}{d-c},$$

去山远 $y=\dfrac{cb}{d-c}$。

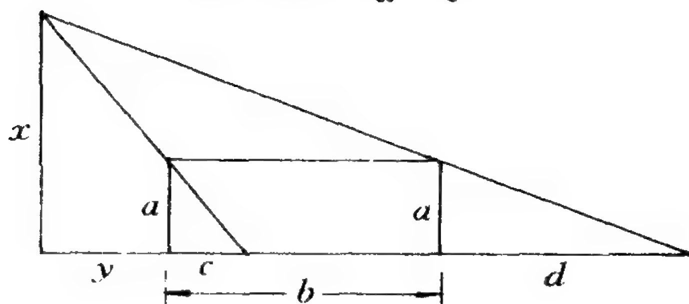


图 8

三上自言不知刘徽用什么方法获得这两公式，但认为可能自下两比例式或与之相当的等式导出：

$$\frac{x}{y+c}=\frac{a}{c},\quad \frac{x}{y+b+d}=\frac{a}{d},$$

并认为“重差”一词意指比例的重复应用，而比例方法在《九章算术》勾股章中多次出现，以为上说佐证。三上还因此说：

“刘徽对这些复杂问题的研究，说明他对代数处理的某些方面颇为娴熟。如果对《九章算术》中某些章节的内容细加推敲，我们就不能不把代数学的诞生，归功于中国在它历史上早期的数学家们。这个时期比秦九韶与李冶出现的 13 世纪要早得多，而一般（按应指西方数学史家）相信秦李之时才是中国代数学诞生的时代。正因为如此，所以我们特别强调刘徽关于这些复杂问题的考虑，它们出现于公元 3 世纪时，在天元术代数学曙光初露之前整整有一千年。”

近代如李俨、钱宝琮等对《海岛算经》诸题都有注释与证明〔乙 1—7〕。但他们的证明基本上沿袭利玛窦、徐光启与李潢之法，添作平行线（与）用相似三角形，或如三上义夫那样应用代数符号列出方程并进行代数运算以求解。有人还甚至应用了三角函数来证明。国外一些学者如李约瑟等也大都用类似方法。但有别于此，特别应该提到的是李俨在《重差术源流及

其新注》〔丙 1〕一文中关于杨辉海岛公式分析的一个介绍。现将李俨此说全文照录如下。(图 9，10)：

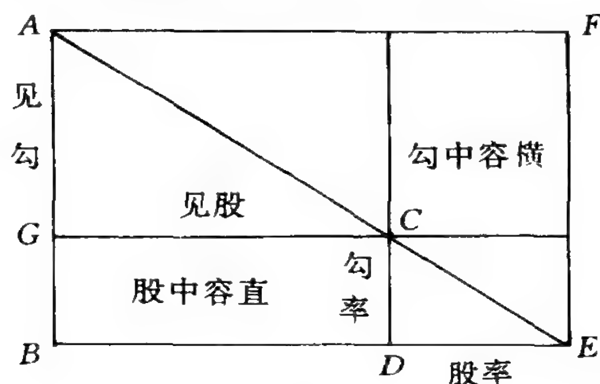


图 9

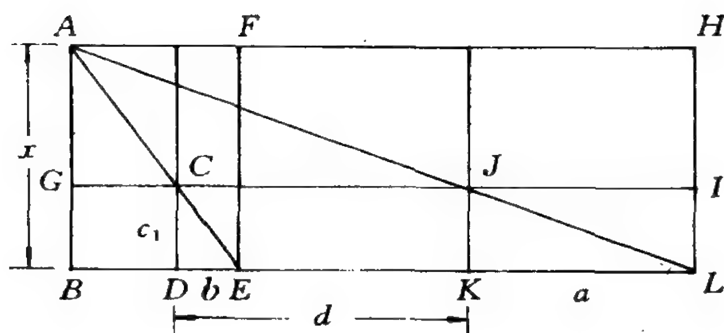


图 10

“其法于 ABE 直角三角形（参阅图 9——编者），知 $AB=$ 勾，$BE=$ 股，$AE=$ 弦；$CD=$ 余勾，$DE=$ 余股，而 AE 弦之内外。分二勾股；其一，勾中容横，如 CF 长方形，其一，股中容直，如 BC 长方形，二积之数皆同，即$\square BC=\square CF$。故

$$AG=\frac{BD\times CD}{DE}。$$

于海岛第一题（参阅图 10——编者），因$\square BC=\square CF$，而 $DE=$ 小余股；又因$\square BJ=\square JH$，而 $KL=$ 大余股，故 $\square BJ-\square BC=\square JH-\square CF$，即$\square DJ=\square JH-\square CF$，即 $CD\times DK=$（$KL-DE$）$\times AG$；$\therefore c_1\times d=$（$a-b$）$\times AG$，

即 $AG=c_1\times d/(a-b)$，而 $x=\dfrac{c_1\times d}{a-b}+c_1$ 矣。

明利玛窦、徐光启译《几何原本》其第一卷第43题称“凡方形对角线旁两余方形自相等”。辉所取者，盖此义也”。

四、对这些论证的评论

上节所列举各家的论证，我们认为除杨辉的论证以及李俨对杨辉的论证所作解释以外，其他则不仅与中国古代几何学的真意不符，说得严厉一些，可以说所举证明都是“错误”的。

上列各家证明除杨、李的论证外可分两大类。其一是添一些平行线或（与）利用相似对应边成比例等定理。其二是把所求高远作为未知数用 x，y 等表示，依相似三角形列出代数方程而求解。

先就第一类证明来说。平行线与相似形理论都是欧几里得几何学中的重要构成部分。这些原始直观的概念自然会以某种形式出现于我国古代几何学中，事实上也确是如此，但显然从未占据象在欧几里得系统中那样特殊的地位。就作者的调查，在我国经典算术中只发现宋秦九韶《数书九章》（1247年）中第六卷田域类漂田堆积题的附图中添有一条平行线。但这在我国古代文献中似是一个例外，且术文无法理解，不能用一般相似三角形边成比例的定理来解释。Libbrecht〔A6〕曾指出后人必有篡改。附图的来历有待查明*。不论如何，我国古代几

---

* 补记　现查知，据宋景昌《数书九章札记》，现存图为后人改作，已非《永乐大典》中原图，平行线系沈钦裴所添入。《札记》引沈氏之言曰：“更增一虚线，与……平行，则比例之理显矣”。这又一次说明古希腊的平行线理论与我国古代几何无缘，那种随便依据欧几里得几何来解释我国古代几何的做法都是极为错误的。

何学自有其独特的体系，从内容到方法都与欧几里得体系异其旨趣。我国从来没有象西方那样在一条还是几条平行线上纠缠不清而另有发展重点，这正好说明我国几何学的特点与其高超之处。

就在同一书《数书九章》中，第七、八两卷测望九问大都与《海岛算经》诸题类似，其中望山高远一题更与望海岛题类同，而附图残缺。宋景昌在读《数书九章》后所作《札记》中依据明《永乐大典》将该图补入。图中并无平行线（图 11）。利玛窦与徐光启在《测量法义》与《测量异同》的证明中也无平行线之设。直到李潢证中才出现一条平行线。可以想见平行线的添入是直到欧几里得几何在我国知识分子中已牢牢扎根的清代才想到使用的。这说明钱宝琮在重绘日高图时所添的一条平行线（图 2 中右下方）远远脱离了我国古代的实际情况。它的添入是毫无根据的。

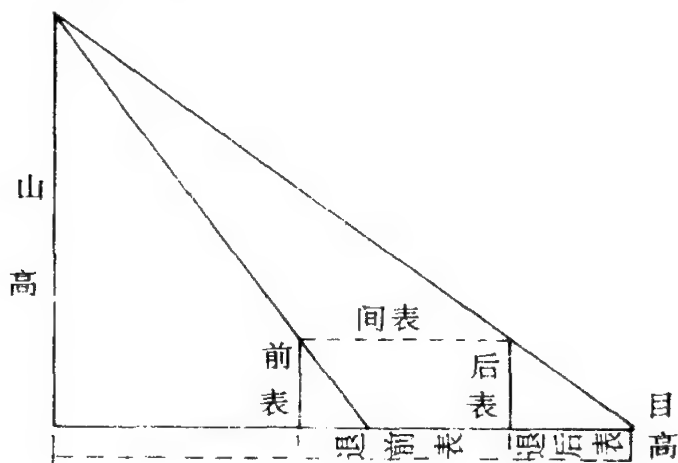


图 11

至于相似形理论，则相似勾股形对应勾股成比例的命题在我国有着悠久的历史。不仅刘徽《九章注》勾股章自第 14 题以下至最后第 24 题的证明中每题都曾用到这一命题，早在《周髀算经》中即已是求日径方法的依据。但在古籍中并未见

到有一般相似三角形的论述。应用一般的相似三角形来证明《海岛算经》诸公式也是缺乏根据的。

因之第一类的证明可以认为是利玛窦等传入西算以后我国自徐光启以下应用西方欧几里得几何所作，而与中国的古代几何学并不相干。这些证明往往要添许多平行线，推理过程也颇繁复而不自然。例如钱宝琮就曾评论过李潢的证明：“图中添线过多，恐不能符合刘徽造术原意。”李潢在《重差图序》中自己也说：“图中以四边形五边形立说，似与勾股不类。”事实上，欧几里得几何学演绎体系的一个根本弱点，是它难以据此以获得新事实。如果事先没有其他方法得到正确结论，即使要证也无从证起。以《海岛算经》诸题公式之复杂以及测量方式之奥妙，如果没有一个深邃的思想作为指导，是很难得出这些公式来的。而这正是欧几里得体系所无能为力的。

第二类代数方法的证明则更一无是处。代数学是我国独特的创造之一，自《九章算术》以至宋元发展线索极为分明。但未知数作为一个独立而明确的概念出现，则是在宋金时代十二、三世纪时，即所谓天元一。从一个未知数到几个未知数、列出方程以及各种代数式的运算规律则一直到元代始集其大成。毛主席曾指出，从认识的感性阶段上升到理性阶段以及新概念的产生，都是一种认识上的飞跃。从《九章算术》的线性联立方程以及某些二次方程的雏型到天元一以至近代的代数符号系统与运算体系要经历过多次这种认识上的飞跃。每一次飞跃都要在大量实践活动的基础上经过相当长的时间往往几十百年之久才能出现。用现在通行的代数学对古代的数学作一些单纯的注释以方便现代读者自然是无可厚非也是既值得做又应该做的工作。但若据以推理，以现代数学来代替一千多年以前的数学，则是绝对不能允许的。

即如以三上义夫对《海岛算经》的解释来说，三上所举：

$$\frac{x}{y+c}=\frac{a}{c}与\frac{x}{y+b+d}=\frac{a}{d}$$

两个比例式，应用了相似勾股形的比例关系。后者屡见于《九章算术》勾股章的刘徽注中，这里 x，y，a，b 等可作为“山高”，“去山远”等等的简写，自然是可以的。但三上从这两式作一些代数运算以导出 x，y 并据以说刘徽已拥有宋元时期的一些代数学知识，则是缺少根据而是违反时代的。三上针对当时许多西方数学史家往往一笔抹煞中国古代数学成就的帝国主义作风，依据他对中国古代数学比较丰富的知识来予以驳斥，这种坚持正义的精神中国人民应给以高度评价。尽管如此，他的论证不仅不符合实际情况，还应认为是“错误”的。

至于用三角函数来证明《海岛算经》诸题则更可不必讨论。

总之，那种用现代代数方法来证明《海岛算经》诸题的尝试在方便现代读者来说自然有其积极意义的一面。但此外则充其量只是用现代方法验证了刘徽的那些公式与定理都正确无误而已。这非特完全不足以反映当时的数学情况，反而使古代辉煌的成就因之而淹没不彰。这正象用 17 世纪以来的微积分这一现代化武器把阿基米德计算抛物线弓形面积的公式重新证明一遍。这在现代自然是轻而易举的事，对稍学过一些微积分者也只是一个简单的练习。但若因此而得出结论说阿基米德的结果与定理没有错，则不但没有什么意义可言，甚至颠倒了历史，反而把阿基米德的重要贡献贬低了。

五、数学史研究中的方法问题

我们所以不惜用大量篇幅列举海岛公式的后代各种证明，指出它们的不当之处，并且强调除个别如杨辉、李俨另当别论

外，这些证明特别是滥用代数符号者都是“错误”的，乃是因为这些错误方法已泛滥于绝大多数流行的数学史著作之中，致使古代数学的真实情况不仅淹没不彰，而且面目全非，许多巴比伦神话、印度神话以及丢番图神话之所以产生，这是重要原因之一。下面是一个典型的例子。

斯特洛伊克的《数学简史》〔A14〕在论及公元前近二千年巴比伦的数学时，举出了一个记载于巴比伦泥板的例子：

由两个正方形并组成一个面积为1000，一正方形边为他正方形边的2/3减10。两正方形的边如何？

根据这一类例子作者就下结论说：

“汉穆拉比时代的巴比伦人已完全具有了处理二次方程的技术。他们能解两个变数的一次和二次方程，甚而至于包括三次和四次方程的问题。”

另一个数学史家Scott在这一问题上他所著《数学史》〔A11〕11页第1行虽然也有其他类似的谬论但态度要严肃些。现把《数学史》中有关的一段译述如下，以与上文对照：

> “巴比伦人曾被认为具有解从两个未知数得出二次方程的方法的知识。这种论断是由于他们企图解决下面的问题而产生：已知长方形的周与面积，求它的边，即解方程组 $x + y = a$，$xy = b$（作者注：这种等同是一种大错误）。这种问题在古代是很普遍的；这可能是为了否决某种广泛流传的信念而引起；这种信念认为一个平面形的面积由它的周界所完全决定，正象在圆与正方形的情形那样：相同周界的图形具有相同的面积。只要这种信念依然存在，丧失道德的商人们就有无限的机会。这个问题曾被希腊人有系统地处理过（作者注：这个说法是不正确的）。至于巴比伦人的方法则是经验的（作者注：重点是作者所

加)。他们的方法是把周界保持不变而把面积所能取得的不同数值确定出来。例如对一个周长为40的长方形来说，结果可列表如下：

$$a+b=20$$

a	b	面积 ab
10	10	100
8	12	96
6	14	84
4	16	64

如此等等。”

我们把两个数学史家著作中的某些片段并列如上，读者们可相互比较，自己作出结论。

古代巴比伦人就象古代许多其他民族一样，从事各种生产活动，观察天象，进行计算，为此付出了辛勤劳动。他们对科学文化所作贡献在我们现代的日常生活中还留下痕迹，留传至今的巴比伦泥板上各种天文数表的记录都是人类伟大的文化遗产。但是，一切科学归根结底都来自生产实践也受着当时生产条件的制约而不能超越时代。知道长方形的周长与面积以求边这一类问题，在古代许多不同民族中本来是会自然产生的。但生产的发展会推动数学的前进，随之而在这类问题的解法上也会显出先后进展的痕迹。试以我国为例。以有书可稽的标准来说，则赵爽《句股圆方图说》包括了二十来个定理，其中最后一个定理即是已知勾股和与勾股积求勾与股的明显公式。在《九章算术》勾股章中更有不少比它复杂得多的问题与解法。但这些解法都是纯几何的，大都依据出入相补原理。自此以后，仍以有书可稽者为准，则杨辉在《田亩比类乘除捷法》卷下（1275年）所引刘益《议古根源》（12世纪）的22问中，有8个都是已知直田长阔和或差以及直田田积要求长阔一类的

问题，所用方法都是增乘开方法。到朱世杰《算学启蒙》(1299年）开方释锁门共34问中，第8至第11四题也是这一类问题，但已进步到使用天元术。据朱世杰原注：谓“按此以古法演之”，除重述了前面所引见于赵爽《句股圆方图说》中的“古法”外，并说：“今以天元演之，明源活法，省功数倍”。在《算学启蒙》(1299）中，朱世杰还只用一个未知数天元一，到了《四元玉鉴》(1303年），则在两仪合辙门的12个同类问题中，朱都列天元地元两个未知数的联立方程组来求解。我国古代算书散失者多而留存者少。但即就现有存书有案可稽者来说，从《句股圆方图》与《九章算术》的几何解法到朱世杰联立方程组的解法，发展线索仍极为分明。其间相隔一千多年，不知要经历多少次认识上的飞跃才能达到。单是从贾宪、刘益的增乘开方到李冶的天元术，再从一个未知数的天元术到两个未知数的天地两元之术，每一步都是极不容易的发明创造。这本来符合马列主义的认识论而是科学发展的客观规律。我们可以说《句股圆方图说》与《九章算术》中的几何解法是导致我国宋元代数的某种源泉与滥觞，起过一定的推动启发作用，但决不能贸然视之为就是代数学。正是同样的理由，尽管我们尊重古代巴比伦人的伟大成就，但是把近代的代数轻易地加之于古代巴比伦人身上，不能不说是十分荒谬的。

六、《海岛算经》的发展与影响

自刘徽以后，虽然史料残缺，但《海岛算经》直到明末仍有踪迹可寻，现根据不完全调查而有书可稽者摭述如下：

天文观察上重差方法的应用散见于历代史书中天文志、律历志以及天文著作中。

《隋志》、《唐志》都提到有《九章重差图》一卷。

唐初王孝通《辑古算经》表称“徽…造重差之法”。

唐初将刘徽《九章注》第十卷重差那一卷单行而改称为《海岛算经》，立于学官，成为《十部算经》的一部。

南北朝祖冲之、北周甄鸾、唐李淳风等都曾为《周髀算经》或（与）《海岛算经》作注。

南北朝何承天、隋刘焯、唐李淳风等对《周髀算经》量日影“句之损益寸千里”以及其后张衡、郑玄等从其说者多有所批判。历代曾为此去全国各地实测晷影，远至交趾地区，尤以唐一行与元郭守敬的规模与影响最大。

唐李淳风注《周髀》，批评“用日以为高下”为“语术相违，是为大失”。李又推广原求日高术到“地有高下”不平的情形，就“表望不同”作“后高前下术”等六术。

《宋史》载有《海岛算经》一卷，与夏翰《新重演议海岛算经》一卷。

《宋史》（卷 462，方技）谓楚衍（11 世纪初）“于《九章算术》…《海岛算经》诸算经尤得其妙。”按早期的高次方程数值解法即增乘开方法的开创者贾宪（11 世纪）是楚衍的弟子。

宋秦九韶《数书九章》（1247 年）第七、八两卷为测望九问，第十六卷有军旅类的望知敌众，其中陡岸测水、表望方城、望敌远近与表望浮图诸题，秦都说“以重差求之”或“重差入之”。而望山高远一题更与《海岛算经》第一题海岛无异，但改《海岛算经》中的人目著地为目高 5 尺。

宋杨辉现存算书中曾三引《海岛算经》，并对海岛一题进行分析（见前节）。从所说“尝置海岛小图于座右，乃见先贤作法之万一”，推想杨辉还见到过刘徽至少是海岛一题的原图。

南宋有《海岛算经》刻本，据杨辉引述现传本与南宋刻本除先后次序可能有不同外内容无异。

元朱世杰《四元玉鉴》（1303 年）中卷勾股测望门八问中第 4 至第 8 依次类同于《海岛算经》中的望海岛、望松、望邑、望谷与望清渊诸题。

明《永乐大典》（1403—1408 年）收有《海岛算经》，现在的传本即从其中辑出。

明吴敬《九章算法比类大全》（1450 年）引用了《海岛算经》题称为“古问”。

明周述学《神道大编历学算会》（1558 年）卷 3 勾股有两题与《四元玉鉴》勾股测望门第 5，第 8 两问题目数据全同，而这两问又与《海岛算经》中的望松与望清渊两题类同。

明程大位《算法统宗》（1592 年）卷九勾股章之末重述了杨辉对海岛一题的分析。

明末利玛窦、徐光启关于测量的译著都有重差一类问题，已见前节。详细分析见下面第八节。

清李潢《重差图序》中说：“望海岛旧有图解”，可知直到 19 世纪之初李潢时还可能见到原来的海岛残图，具见前节。

刘徽《海岛算经》的原来理论与证明可能在唐宋时期即已鲜有人知而失传，这从杨辉据图以求“源”可以推知一二。但《海岛算经》却在另一方面对我国的数学起着巨大的推动作用并影响及于后世近代数学的发展。

宋元之世我国数学有着多方面的重大成就，天元术的创立是其中之一。一般认为高次代数方程数值解法是宋元时期数学发展高潮的顶峰。国外对我国古代数学所推崇的主要是祖冲之圆周率以及这一成就。但是，我们认为与天元术相伴而生的几何代数化、代数方法在几何上的应用以及代数式的使用与代数运算的系统化等要比这一成就意义重大得多。而重差理论在这方面所起的作用是不小的。检查一下天元术与重差理论的有关

论述可以为它们的相互关系提供一些线索。

首先是秦九韶的《数书九章》。按秦著中测望九问与望知敌众十个问题都有术无法，不仅立术不明，抑且谬误极多。有几题虽提到用重差求之，实则与重差术并无关系。这些都说明秦本人对勾股之术不甚娴熟，对重差之术更不理解，而且疏忽大意，但不能因这些瑕疵而抹煞其主要功绩，按测望问题在秦全书中所占份量颇多，在《海岛算经》之后我国算书中还是第一次作这种较有系统的收辑，足见秦的重视。测望九问之中，有四问都用列高次代数方程求解。虽然据钱宝琮意见，秦对天元术似乎并不了解，但从《九章算术》勾股章第20出方邑望木题的开带从平方历经唐王孝通开带从立方与宋贾宪增乘开方一脉相承，应无疑为天元术的前身，虽无天元之名而有天元之实。秦之测望九问显示出与解高次方程以及天元术的密切关系，秦的这一功绩应予肯定。

至于朱世杰《四元玉鉴》勾股测望八问则全部用天元术求解。在八问中第五问以下都与《海岛算经》诸题类同，其他三问都与勾股，也与秦的某些问题相似。这些问题显系从《九章算术》勾股章出方城诸题推衍而来。只是有时改方城为圆城问题要难得多，非单纯的勾股术所能解，因而也促进了天元术与一般代数运算方法的发展。东汉末郑玄引郑众说"今有重差句股"。刘徽也说"辄造重差…缀于句股之下"。重差与勾股本来是密切相关的。在朱之前早在12世纪之中，勾股问题就已成为天元术重要的问题来源。例如在杨辉《田亩比类乘除捷法》(1275年)一书中编入了已失传的刘益所著《议古根源》(12世纪)二百问中的二十二问，其中有两问都是用天元术求解的勾股问题。李冶的《测圆海镜》(1248年)与《益古演段》(1259年)则全都是以勾股为主题的天元术应用。由此可见勾股曾为天元术提供了许多具体的问题实例。重差术估计也起着

同样的作用。现将朱世杰《四元玉鉴》勾股测望门第4题望圆城也即《海岛算经》的望海岛题朱氏的术与草照录如下（略去求表去城部分），以与海岛术相比较：

朱世杰原文	现代语言译文
术曰：	解法：
立天元一为城高	置城高 $=x$
如积求之，得一千四百四十为正实，四十为益方	列方程得 $-40x+1440=0$
上实下法而一，得高	解之得 $x=\frac{1440}{40}=36$ 尺
草曰：	
立天元一为城高	置城高 $=x$
又为大句	大勾 $=x$
以两表退行相减，余四十尺，为两表退行差	两表退行差 = 后表退行 − 前表退行 = 40 尺
又为小股	小股 = 40 尺
用乘大句，得 太 ||||0 寄左	左 = 大勾 × 小股 $=40x$
乃以两表退行差加表间得一百二十尺为大股	大股 = 两表退行差 + 表间 = 120 尺
以表高为小句	小勾 = 表高 12 尺
乘之，得一千四百四十为同数	大股 × 小勾 = 1440 与左相同
消左得 ||||0 1 ☰ ||||0	移项整理得 $40x-1440=0$
上实下法而一，得三丈六尺为城高	∴ 城高 $x=\frac{1440}{40}=36$ 尺

《四元玉鉴》其他诸题解法仿此。据此我们认为重差与勾股一样，对天元术的发展都不仅有刺激推动作用，甚至还是重

要的源泉。而且，重差诸题的解法只涉及一次方程即所谓开无偶平方，实质上又与刘徽的原证相同（详见下节的分析），只是改用了天元术的语言而已。相反勾股一类问题所引出的方程往往是高次或（与）多元的，要复杂得多。因而重差之成为天元术实例，实际上更容易更简单，出现得也可能较早。只是数书大都失传，作为《四元玉鉴》前身而为祖颐为朱所作序中曾经提到的《益古演段》、《照胆》、《钤经》、《如积释锁》、《〈如积释锁〉细草》、《两仪群英集臻》、《乾坤括囊》等书内容都已无法考查，不能追踪重差与天元术的关系而为之确证而已。

三上义夫曾认为刘徽已能用代数方法证《海岛算经》诸题，我们指出过这个说法是不成立的。三上以及其他人士例如李俨等在用代数方法证明时，都列出两个未知数的方程组求解。从现代的知识说来这种做法是颇为自然的。然而朱世杰《四元玉鉴》中的天元术解法却偏偏只用一个未知数——天元，虽然《四元玉鉴》中用天元、地元两个未知数求解的问题有三十六问之多。由此也可证明非但这些代数证法远不是刘徽原证，甚至还不是天元术出现时应用于《海岛算经》诸题的证法。与之相反，我们认为正是刘徽的《海岛算经》诸题曾为天元术提供了具体实例并促使其发生发展，其作用可能不在勾股一类问题之下。

17 世纪解析几何与微积分的创立是数学发展史上的大事。解析几何的诞生一般认为主要应归功于笛卡儿 Descartes。但是笛卡儿一无坐标系统，二无正负线段，他的主要功绩，不少数学史家认为是在于代数的几何应用与几何的代数化，在于把几何图形表达成代数形式，因而为研究物理世界（笛卡儿认为可归结为几何学）提供了数量工具。有人并因此提出“代数几何”应是一个比解析几何更合适的名词。以研究笛卡儿著名的 Scott 认为笛卡儿的成就是：把一个几何问题化成代数形式因

而得以使用整个代数机器以求解。然而，这种几何与代数的配合、代数的几何应用与几何的代数化正是宋元天元术的主要含义之一。而且就在宋元数学家的手里为了发展天元术而建立了一整套的代数机器，包括天、地、人、物等元的正负乘幂及其代数式，以及这些代数式的运算系统的建立。天元术以勾股重差一类问题作为立术的主要来源，以天、地、人、物等元代替所求线段，用它们的代数式来表示几何图形的长度、面积，然后运用相伴发展了的那套代数机器进行求解。所有这些，正是所谓笛卡儿的最主要的贡献所在，是解析几何得以创立的决定性的一步。当然，笛卡儿在变数的引入以及就在解析几何上的其他种种发明创造对于数学的发展有着划时代的功绩，他对世界数学的伟大贡献是完全肯定的。然而无视于我国天元术在这一方面所起的重要作用，而单纯强调代数方程数值解的成就（这一成就的辉煌是无庸置疑的），把它看成是我国数学的“顶峰”，则实质上无异是贬低了我国对世界数学的伟大贡献。同样，我们认为如果把《海岛算经》的功绩局限于测望之学，而忽略它对天元术发展的影响与作用，则对它的估价也还是不够的。

七、海岛公式证明的复原

我们认为，与其他诸家不同，杨辉对海岛公式的分析以及李俨根据杨辉补出的证明是基本上正确的。

先考虑“句中容横与股中容直二积皆同”这一结论（图9），这是整个重差理论的基础，也就是杨辉所说的“源”。它的来历可能有二。其一是根据相似勾股形的命题：股率乘见勾，等于勾率乘见股。这一命题在《周髀算经》求日径术中即已用到，而见于《九章算术》句股题15以下刘徽注的每一题中。另一是出入相补原理。这是我国古代几何学中面积体积理

论的结晶。所谓“出入相补，各从其类”这一基本原理，明确叙述于勾股章开首勾股术的刘徽注之中，而被广泛应用于《九章算术》的勾股商功等章。

在图 9 的情形，出入相补原理与相似勾股形命题自然是等价的。这在《九章算术》勾股章的徽注中也可看出。按勾股章共 24 题，自第 1 题至第 13 题，徽注的证明都用出入相补，第 17 题以下都用相似勾股形，此外数题则都两者并用。第 15 题勾中容方问题，更详述两种证法，而勾股章自 17 题以下求邑方与量木高井深诸术，在方法上显然都从勾中容方问题演化而来。两种方法在此是相通的，而出入相补原理是一个更高更普遍的概括，这一普遍原理不仅是我国古代几何学的一个结晶，且其影响及于整个数学的其他不同领域。由于牵涉面较广，非简单篇幅所能尽，我们将留待其他机会再作详细讨论。就本文当前的问题来说，则不论属于何种情形，杨辉证明海岛公式的“源”应系直接来自对“海岛小图”的观察与对我国古代遗产的具体分析。李俨所谓“辉所取者，盖此义也”，牵扯到欧几里得原本之说，是没有任何根据的。

但是李俨根据杨辉分析补作的证明，则我们认为是大体上正确的，可以认为实质上这就是刘徽的原证。为说明这一点，我们把证明重述如下，并附以赵爽日高图说中的相应语句，作为此证即是原证之说的佐证（图 12）。

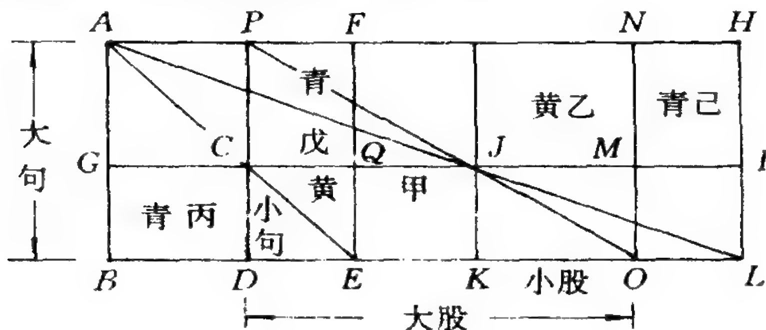


图 12

证　　明	赵爽日高图说
$\square JD=\square JN$，	黄甲与黄乙其实正等，
$\square JB=\square JH$，即	黄甲与青丙相连，黄乙与
$\square JD+\square CB=\square JN$	青己相连，其实亦等。
$+\square MH$，	青丙与青己其实亦等。
$\therefore\square CB=\square MH$。	以表高乘两表相去为黄甲
$\square JD=CD\times DK$，	之实
$\square CB=\square CF$，	
$\therefore\square MH=\square CF$，	（出入相补原理）
由此得 $MI=CQ$。	
$JM=JI-CQ=KL-DE$，	以影差为黄乙之广，
$\therefore\quad MN=\dfrac{CD\times DK}{KL-DE}$，	而一，所得则变为黄乙之袤。
$\therefore$　日高 AB	
$=\dfrac{CD\times DK}{KL-DE}+BG$	上与日齐，按图当加表高。
$=\dfrac{\text{表高}\times\text{表间}}{\text{景差}}+\text{表高}$。	

去表远公式的证明如下：

由青丙＝青戊（出入相补原理），得

去表远　$BD=\dfrac{AG\times\text{前去表}}{\text{表高}}$。

前已知 $AG=\dfrac{\text{表高}\times\text{表间}}{\text{景差}}$，故得　$\text{去表远}=\dfrac{\text{前去表}\times\text{表间}}{\text{景差}}$。

对于上面补出的原证再作一些补充说明如下：

（1）从证明步骤看，现存赵爽《日高图说》似既有脱漏又有句文倒置，从现存甄鸾的注与残图（虽然错误极多）看来，现在所补的证与图似尚未能与原文完全相符（例如有甲乙丙与戊己而无丁）。

（2）证明还可以简单一些，例如 PO、ON 线即可略去，

▭JH 也不必分成黄乙与青己两部分，写成上面较繁的形式是依据《日高图说》的术文，而朱世杰天元术文又可为之佐证，见下（3）。

（3）与朱世杰天元术文（见第六节）比较，可看出朱文（图 12）中

大勾 $= AB$，小股 $= KO$，大股 $= DO$，小勾 $= CD$。故大勾 × 小股 = 大股 × 小勾，即 ▭KN = ▭DM 或相当于黄甲 = 黄乙（都根据出入相补原理），朱世杰的解术无疑是有来历的，在补出原证中添上 PO 与 ON 等线者与此正相符合。

（4）与钱宝琮补出日高图（图 2）比较，可见添作平行线之非理，也可见黄乙与青己的位置应左右对易。（参看下节）

（5）为进一步说明，补出《海岛算经》第二题望松的证明如下（题与术文从略）：

附图 13 中 AB = 松高，CD、FG 为前后两表，CE = 入表，DI、GK 为前后两表退行，依据望海岛题表去山公式，可知视 ED、HG 为望松本 B 点的前后两表，而与 CD、FG 为望松顶 A 两表比较时，LF、MH 应同交于平地上 J 点，使 GJ = 相多，而 AF、BH 应同交于后退表处 K 点，使 JK = 前表却行 DI。今据出入相补，有▭FD = ▭FP，▭HD = ▭HQ。相减得▭EF = ▭NP − ▭HR，或▭NP = ▭EF + ▭HR，即松高 × 相多 = 表间 × 入表 + 入表 × 相多，也即《海岛算经》的松高公式：

$$\text{松高} = \frac{\text{表间} + \text{入表}}{\text{相多}} + \text{入表}。$$

朱世杰《四元玉鉴》勾股测望门第五题与《海岛算经》望松题类同，但易松生山上为方城上戍柚，朱术立天元一为柚高，又为见勾差即图中 AB。相多即 GJ 为股差率，以表间加股差率即 DJ 为见股差，入表即 CE 为勾差率，从见勾差 × 股

差率＝见股差×勾差率也即前面的▭NP＝▭ER。得方程解之得柚高。由此解法也可为所补原证中添作诸线的一个旁证。

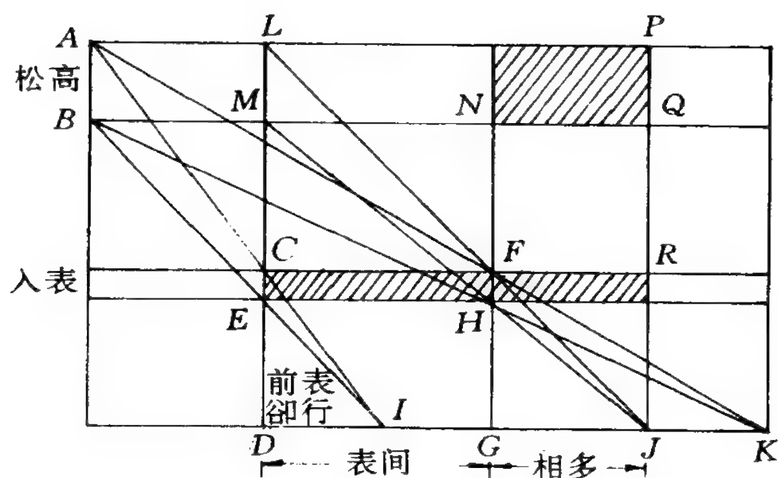


图 13

(6) 今再补出《海岛算经》第四题望谷的证明如下：

图 14 中 AB＝深谷，CD、EF 为岸上重矩之勾，CG、

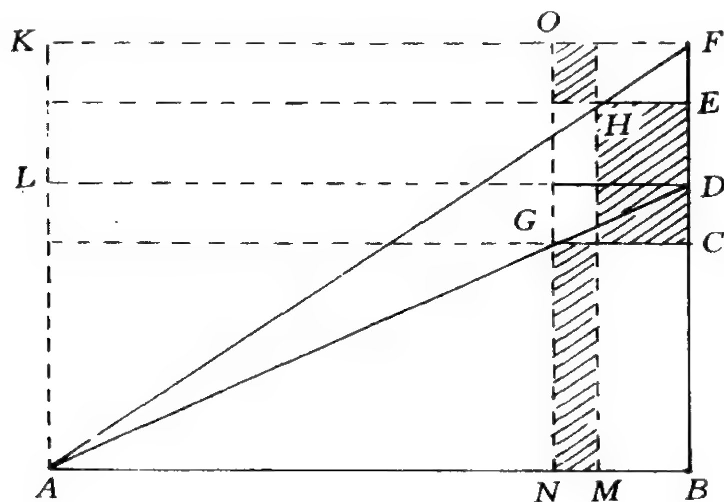


图 14

EH 为测谷深处 A 点所得入股。依出入相补得▭HK＝▭HB，▭GL＝▭GB。相减得▭HO＝▭HC－▭GM，或▭GM＝▭HC－▭HO，由此即得《海岛算经》中的谷深公式：

$$谷深 = \frac{矩间 \times 上股}{下股 - 上股} - 勾高。$$

朱世杰《四元玉鉴》勾股测望门第七题与《海岛算经》望谷题类同，所用天元术草亦与《海岛算经》术文相同，只是用天元术语言表达而已。

(7) 朱世杰《四元玉鉴》勾股测望门第六第八两问即《海岛算经》中望邑与望清渊两题，其天元术草也与《海岛算经》术文一致。朱书中这几问的方程都是一次的，即所谓开无偶平方。这些都足以说明一方面朱书这一部分渊源于《海岛算经》原著，另方面《海岛算经》为天元术提供了最简单最直接的素材，因而应是天元术产生的早期的推动力。

(8)《海岛算经》其他诸题，都易从海岛公式以及松高谷深等公式推出。这些证明都简单直观，容易为一般人所接受，这正是我国古代几何学的一个特色。且证明直接导致《海岛算经》诸问原来给出公式的形式，与之相反，不论如李潢用欧几里得几何或后世诸人如李俨〔乙 2〕用近代的代数，证明都颇繁复，尤其要再经历一番曲折与演算，才能凑出《海岛算经》原来的公式形式。这也说明这些证法与刘徽原意毫无共同之处，而上面所提出的证明，才是符合刘徽原意的证明。

(9) 从李潢《重差图序》，知李曾见到海岛旧图，且所谓“两形相去辽远”而须“宛转通之”，正是出入相补的特色，因而可以相信这一旧图即如图 12 那样。至于所谓“皆可不必也”之语，如果把李潢《海岛算经细草图说》与我们补作的图与证明相对照，优劣判然。这正好说明明末以来中国固有的几何学已为欧几里得几何所排挤殆尽，清代学者如李潢几已一无所知，只能屈从于欧几里得的一统天下而已。

丢开对天元术所起的推动作用不谈，只从测望这一角度来说，《海岛算经》立意造术之深，也非西方如海伦等所能企及。

而且，《海岛算经》似已发展成一个完整的系统理论。依据刘徽《九章算术注》原序中所说："度高者重表，测深者累矩，孤离者三望，离而又旁求者四望，触类而长之，则虽幽遐诡伏，靡所不入，博物君子，详而览焉"。似乎刘徽已总结出一般的规律，不论地形如何"幽遐诡伏"，都可以设计出何处设表，作出测量的具体方案来。刘徽原意虽然已不得而知，但这对于现代的数学工作者来说，仍是一个值得思考与解决的问题。而且，重差理论中以量长代替角的测量这一方法所隐含的以多次简易测量代替较难测量的原理在现代的各种技术问题上可能还是有现实意义的。

八、利玛窦来华时期的测望理论及其影响

综观我国测望理论的历史演变过程，明代以至利玛窦、徐光启时期可以说是一个转折点。在第六节中已经指出，我国古代测望理论从《周髀算经》时起，以迄宋元之世不仅一直在发展，而且在推动天元术的发展因而对近代代数学与解析几何的影响是无可估量的。到了明朝八股取士，理学统治了学术界的思想，数学已趋衰微，但《海岛算经》仍有传者。到利玛窦于1582年来华，1600年进入明室宫廷，士大夫如徐光启、李之藻争相结纳，推崇西学而对于自己多所诋毁，更使我国古代传统数学一蹶不振，《海岛算经》理论也竟成绝学。

公元1607年（明神宗万历35年）利玛窦与徐光启共译《几何原本》，引入了古希腊的欧几里得几何学。1607至1608年间，又由利玛窦口译，徐光启笔受《测量法义》一书，把西方的测量技术传入了我国。分析一下《测量法义》一书的内容与方法，对于了解它的意义是必要的。

《测量法义》书首"最目"云：

“先造器，次论景，本题 15 首，附三数算法”。

“造器”明确测量目的与所用仪器：“测量者，以测望知山岳柚台之高、井谷之深、土田道里之远近也。其法先造一测望之器，名曰矩度…”，

“论景”指出测量方法：

“法中俱用直景倒景布算，故先正解二景之义，次解其转合于矩度，以资后论…”。

其次是“本题 15 首”，列举各种具体问题的测量方法。最后的附录“三数算法”系比例算法，则即《九章算术》中的今有术。

《测量法义》中所用仪器以矩度为主，15 题中，除第 9 题用平镜，第 10，11，14 用表，第 12 题用矩尺外，其余，都用矩度。矩度、矩尺与平镜特别是矩度是当时欧洲所用主要的测量仪器。矩度与矩尺系 16 世纪中意大利 Tartaglia（因解三次方程而为后世所知）为炮兵发射炮弹而设计〔*B*〕，其中附图与《测量法义》中矩度图完全相同。至于表则为我国自古所用的传统测量仪器，而不见于西方。例如在《测量法义》之后耶稣会教士罗雅谷、汤若望所著《测量全义》一书仪器图说部分曾列举西方古今所用各种测量仪器，即无表的名目。Smith 一书〔*A*13〕几何工具一章论之甚详，也无表的使用，只说古曾有之而已。

Gunther 一书〔*A*5〕有一章专论古代测望仪器，也只说古代有表，现已不用。大抵以表测量是古代许多民族所共同有过的实践活动，但只有我国才发展到重差术这样的理论高度。这正显示出我国在数学上独特的创造力，为世界其他任何民族所不能企及的。

就测量方法而论，则《测量法义》15 题中，除第 10 题以表测高用表两测，第 15 题用表四测，以及第 6 题以目测高的

后面部分用矩度两测外，其余都用 1 测。其中两测以上诸题都是我国古代的经典成果，在第二节末早已指出，重测正是重差理论的指导思想，徐光启继《测量法义》之后作《测量异同》，序言中称：

> "九章算法，句股篇中故有用表用矩尺测量数条〔作者注：这里的矩尺与利玛窦书中的矩尺完全不同〕，与今译《测量法义》相较，其法略同…。既具新论，以考旧义…今悉存诸法，对题胪列，惟求同异…共为测量异同六首…"。

在徐列举异同六首之中，第 5 首以四表测远，与《测量法义》第 14 题以四表测远类同，徐也指出"此旧法与今译同"。实则此题即《九章算术》勾股章的第 22 题立四表求木去人。又《测量法义》中的第 10 题以表测高，则与《测量异同》中第 4 首以重表兼测无远之高无高之远与第 6 首以重矩兼测无广之深无深之广类同，实则与《海岛》第一题无异，只是《测量法义》与《测量异同》中的证明方法都依据欧几里得《几何原本》而已（见第三节），《测量法义》第 4 题的后面部分则无非将原来用表两测者改为用矩度两测而已。

徐光启在《测量异同》序言中只提到九章算法勾股篇，但异同六首既都是"旧义"而"悉存诸法"或"旧篇所有"，则第 4 第 6 两首必有所本。可见徐至少应见到《海岛算经》的第一题，鉴于《海岛算经》明时已收罗于《永乐大典》，今本《海岛算经》即从其中抽出，徐似不应该不见到该书。然而徐在《题测量法义》中却说：

> "刘徽、沈存中（括）之流，皆尝言测望矣，能说一表，不能说重表也"。

这是难以自圆其说的。徐光启指摘明代"算数之学特废"的原因是"名理之儒土苴天下之实事"与"妖妄之术谬言数有神

理"，又提出对历法应"详加测验"、"不差不改，不验不用"，徐又撰《农政全书》，这是我国古代的一部农业百科全书，这些见解与工作都应肯定，但不能因此而为徐对西算的投降主义态度辩解。

至于利玛窦，则史载于1582年（明万历10年）到广东中山，1589年到南京时与徐光启相识，1606年以来与徐共译《几何原本》与译授《测量法义》时，来我国已二十余年。在这样长的时期内说利不接触到中国古代的固有数学，将是不可思议的。西方的测量技术本极落后，文艺复兴以前，西方关于测高望远之学主要见于海伦的《On the dioptra》一书。按海伦的年代从公元前150年至公元后250年估计大有出入，甚至有人与10世纪时拜占庭的某一海伦相混，所著书最早在1814年用意大利文编集刊行，其中内容已知有他人掺入痕迹。所用测望仪器Dioptra大致是一种拟想中物，实际上可能并不存在。即使丢开这些不谈，海伦一书的内容据柯恩（Cohen）及笛卡儿（D escartes）所著书〔A3，6〕所引也不能与《海岛算经》之有指导思想并成为系统理论者相提并论。到16世纪时Tartaglia制作矩度，成为欧洲测望的主要仪器。但器颇笨重，使用并不灵便，在耿受（Gunther）及斯密思（Smith）所著书〔A5，13〕中有不少图片可见其测法之一班。《测量法义》主要部分正反映了当时欧洲的这一水平。而其中较先进的诸题恰与《九章算术》勾股以及《海岛算经》相重，所用表测更与通行西器及其测量方式有别，这说明它们应另有来历。特别是在利玛窦以表测高两测部分的证明中（见前第三节以及图6），须在丑庚线上取卯点使　已卯＝丑寅。按这样的取法正与我国古代海岛公式原证中黄乙在青已左方的位置相当（图12）。如果依欧几里得体系的证法则应添作平行线如图7，这样则所取卯点应使辛卯＝丑寅，或相当于钱宝琮日高图（图2）中黄乙

在青己右方位置的情形。但利玛窦不以此之图（对他说来这应是很自然的），却毫无理由不伦不类地选取了使己卯＝丑寅的卯点，这正说明利的这一问题与证法有它不同于西方欧几里得体系的来源，为了要纳入欧几里得体系，因而硬凑答数留下了斧凿痕迹。至于利玛窦证明之兼形式主义繁琐哲学与矫揉做作三者而有之，则犹其余事也。

在同一时期，由利玛窦授李之藻演的《同文算指》（1614年）中卷之四叠借互征法为利玛窦所据以翻译的 Clavius 即所谓丁先生原著所无，而显然取意于《九章算术》方程章或我国以后算书中的相类篇章。这是我国古代数学的辉煌成就之一，与《同文算指》其他部份相较，水平高下极为显著。然而，李之藻在序中虽自谓“间取九章补缀”，但都成为“利玛窦授”的材料，书首仅题“李之藻演”而已。同样，后来的《测量全书》中也提到阳马、鳖臑、堑堵等《九章算术》中的专有名词而不及《九章算术》。说这些我国古代的文化遗产，不因利玛窦等传教士们的“传授”而流入西方，也将是不可思议的事。本人见闻有限，只能略举数例。在这一西方科学技术开始大规模输入期间，东西方交流中双方的得失如何以及存在的许多问题值得来一番细致调查把它彻底弄清。

自利、徐输入西算以来，终清代两百余年我国对数学的发展已丧失推动作用，自此以来，人人知有欧几里得，而对我国实则更为辉煌的古代几何学却缺少认识。清代中叶虽对我国传统数学重新发掘并有所重视，但如李潢等局限于以欧几里得几何等西算铨释我国古算之类，系统既然不同，自然格格不入，至于所谓创造性的工作，则远落人后，在清一代，一些最有成就者如梅文鼎等，志趣不可谓不高，用力不可谓不勤，也正如钱宝琮在《中国数学史》中所说：“从其具体的数学成果讲来，大都较西欧数学的同样成果迟了一百余年”。这与宋元以前我

国数学比之西方在许多方面遥遥领先者正可对照。

就测量技术而言，沈康身同志在“我国古代测量技术的成就”〔丙 2〕一文中曾有很好的介绍。我们将借用沈文的结语（也适用于测望理论）照录如下，以作本文的结束：

> “明末外国传教士随着宗教把某些科学技术进入中国。当时士大夫阶层，一方面脱离劳动人民，一方面又不认真研习数学。于是崇外之风从此孳生。当时象徐光启这样传播西学的代表人物，在序西海利玛窦授李之藻演的《同文算指》（1614 年）里就说：“观利公与同事诸先生所言历法诸事，即其数学精妙，比之于汉唐之世十百倍之……振之（李之藻）两度居燕，译得其算术若干卷。既脱稿，余始问请共读之，共讲之。大率与旧术同者，旧所弗及也；与旧术异者，则旧所未之有也。旧术与西术合者，靡弗与理合也；与西术谬者，靡弗与理谬也”。他最后的结论是，“虽失十经，如弃敝屣矣”。措辞推重外国，对我国固有创造发明反多诋毁，数典忘祖，这是极不应该的”。

参 考 文 献

甲

[1]《算经十书》，钱玉琮校点，中华书局，1963。

[2]《周髀算经》。

[3] 魏刘徽，《注九章算术》。

[4] 魏刘徽，《海岛算经》。

[5] 宋秦九韶，《数书九章》，1247。

[6]$_1$ 宋杨辉，《续古摘奇算法》，1275。

[6]$_2$ 宋杨辉，《田亩比类乘除捷法》，1275。

[7]$_1$ 元朱世杰，《四元玉鉴》，1303。

[7]$_2$ 元朱世杰，《算学启蒙》，1299。

[8] 明吴敬，《九章算法比类大全》，1450。

[9] 明周述学，《神道大编历学算会》，1558。
[10] 明程大位，《算法统宗》，1592。
[11] 利玛窦口译，徐光启笔受，《测量法义》，1607—8。
[12] 明徐光启，《测量异同》。
[13] 利玛窦授，李之藻演，《同文算指》，1614。
[14] 罗雅谷、汤若望撰订，《测量全书》。
[15] 清李潢，《海岛算经细草图说》，1820。
[16] 清宋景昌，《数书九章札记》，1842。
乙
[1] 李俨，《中国数学大纲》，科学出版社，1958。
[2] 李俨，《中国古代数学史料》第26章海岛算经新注，中国科学图书公司，1954。
[3] 李俨，《中国算学史》，商务，1936。
[4] 李俨、杜石然，《中国古代数学简史》，中华，1963—4。
[5] 钱宝琮，《中国数学史》，科学出版社，1964。
[6] 钱宝琮，《中国数学史话》，中国青年出版社，1957。
[7] 许莼舫，《三角术的嚆矢》，见《中算家的几何学研究》，第2章，1952。
丙
[1] 李俨，《重差术源流及其新注》，见《学艺》7卷8期，1926，1—15。
[2] 沈康身，《我国古代测量技术的成就》，科学史集刊8，1965，28—41。
[3] 张家泰，《登封观星台和元初天文观察的成就》，见《考古》，1976第2期，95～102。
[4] 钱宝琮，《中国古代数学的伟大成就》，见《科学通报》2卷10期，1951，1041—1043。
[5] 钱宝琮，《秦九韶〈数书九章〉研究》，见《宋元数学史论文集》，科学出版社（1966）60—103。
[6] 梅荣昭，《李冶及其数学著作》，同上，149—165。
[7] 严敦杰，《宋杨辉算书考》，同上，149—165。

[8] 杜石然，《朱世杰研究》，同上，166—209。

[9] 白尚恕，《秦九韶测望九问造术之探讨》，同上，附录 4，290—303。

[10] 白尚恕，《刘徽海岛算经造术之探讨》。

A.

[1] Ch. Boyer, History of analytic Geometry, 1956.

[2] F. Cajori, A history of mathematics, 2nd ed, 1919.

[3] M. R. Cohen - I, E. Drabkon, A, source book in Greek science. 1948.

[4] R. Descartes, La géométrie, Paris, 1927,

[5] R. Gun ther, Early science in Oxford, v. 1., Pt, Ⅲ - Ⅳ, physics and Surveying, 1942.

[6] T. Heath, A history of Greek mathematics, 2v., 1912.

[7] M, Kline, Mathematical thought from ancient to modern time, 1972.

[8] V. Libbrecht, Chinese mathematics in the 13th century, 1973.

[9] Y. Mikami, The development of mathematics in China and Japan, 1913.

[10] J. Needham, Science and civilization in China, V. 3. 1959.

[11] J. F. Scott, A history of mathematics, 1958.

[12] J. F. Scott, The scientific work of Rone Descarteé, 1952.

[13] D. E. Smith, History of mathematics, vl, 1923; v2, 1925.

[14] D. J. Struik, A concise history of mathematics, 1948. (关娴译, 数学简史 1956)

B.

[1] Drake, Galileo and first mechanical computing device, Scientific American, 1976 April, 104—113.

《海岛算经》古证探源*

刘徽《海岛算经》是一部关于测高望远之术的专著。按刘徽自序，有"析理以辞，解体用图"，以及"辄造重差，并为注解"等语，说明原著应有注有图。传至后世，所载九题只有方法结果而无注析附图，杨辉在《算法通变本末》卷上中曾说，"海岛算法，隐奥莫得其秘"，清代李潢、沈钦裴等都曾尝试补出证明，但与刘徽原意显无共同之处。我们认为，要使古证复原，应该遵循以下三项原则，是否妥当，还请有志者共同商讨。

原则之一，证明应符合当时本地区数学发展的实际情况，而不能套用现代的或其他地区的数学成果与方法。

原则之二，证明应有史实史料上的依据，不能凭空臆造。

原则之三，证明应自然地导致所求证的结果或公式，而不应为了达到预知结果以致出现不合情理的人为雕琢痕迹。

《海岛算经》的第一题海岛公式从明清以迄近代出现过不少证明，是否合于古证原意，可以依据这些原则逐一考核，中外经典著作，流传至今，历代难免有后人随意篡改增添，甚至伪造伪托，以至真假不辨，也可以依据这些原则作为鉴别真伪的一种手段。

本文将依据这些原则，尝试将《海岛算经》九题的古证补出于下。中国古代的几何学有其自己的发展过程与方法体系，与希腊欧几里得几何迥乎不同。在我国古代几何中，并未见到明

---

* 本文摘自《九章算术与刘徽》(吴文俊主编)，北京师范大学出版社，1982 年。

显的平行线概念，角度也很少用，虽有比例理论，但文献所载局限于勾股相似形的简单比例关系，而几乎未见到有关一般相似形以及各种比例式的变形转化等记载，在古证复原中，这些概念都应避免使用，象李潢[9]那样滥添平行线的作法，更难允许。

相反，出入相补，各从其类的原理，明显见诸九章刘徽注以及赵爽著作，在《九章算术》的商功句股诸章曾多方面应用，由此原理可得以下推论：

推论之一 如图1从▭AB与对角线AB上一点C可得▭CD=▭CE(=指面积相等)。

推论之二 如图2，在句股相似形ADF、ABC与BDE中，称AC=见股，BC=见句，BE=股率，DE=句率，AF=大股，DF=大句(或大股之句)，则有

$$\frac{大股}{大句}=\frac{见股}{见句}=\frac{股率}{句率}。$$

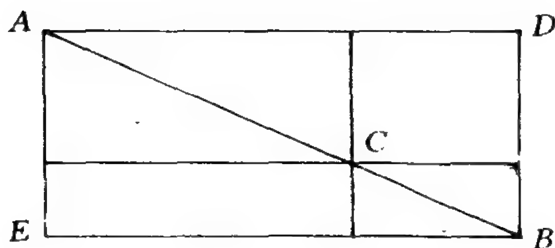


图1

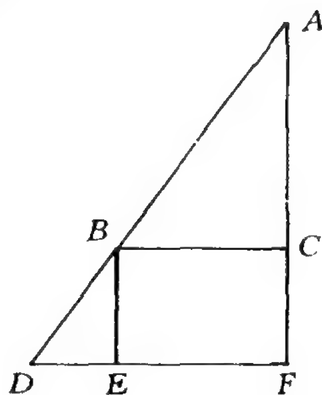


图2

这两推论显然是等价的，都曾多次应用于各种勾股问题，详见刘徽《九章注》以及赵爽《日高图说》与《句股圆方图说》。

本文关于《海岛》九题的复原古证，无非是这两推论的反复运用。九题中海岛第一题中的岛高公式，望松第二题的松高辅助公式(见下)，以及望谷第四题的谷深公式，代表了古代用矩立表以望高、知远、测深的三个基本结果，其余诸题的公式皆可从

这三个基本公式容易得出。在这三个基本公式特别是岛高公式的证明中，我们都只用了推论之一的矩形等积关系而避免应用推论之二的勾股相似形比例关系。其理由可概述如下：

1．这一证明及其附图(图 3)，有赵爽的《日高图说》以及流传至今的日高残图作为依据。所谓“上与日齐”，即说明原图应补成若干长方形。从《日高图说》，可以推想古日高图(图 3)与古证应较本文所示稍复杂一些，详见[14]。

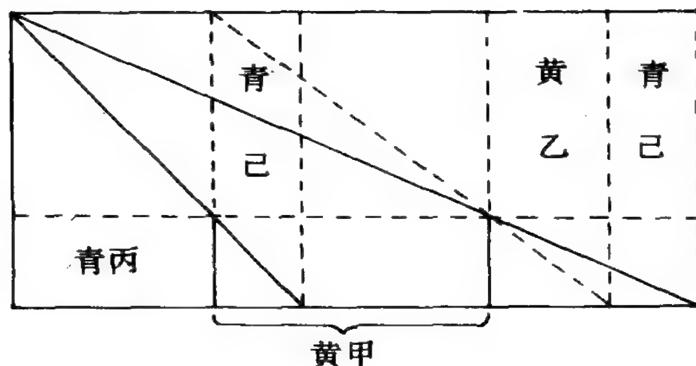


图 3

2．杨辉在《续古摘奇算法》中自言“辉尝置海岛小图于座右，乃见先贤作法之万一”。可知杨辉是依据当时仍在流传的古海岛图来论证的。本文所附海岛图(图 4)，相当于杨辉的隔水量竿的附图，证明也实质上即杨辉的证明或古时的海岛原证。

3．李潢在《重差图序》中说“望海岛旧有图解”，又说“旧图于形外别作同积二方”。这说明李潢时还能见到海岛古图(依李序即古重差图)，而从“同积二方”一语看来，这古图中是有许多长方形从形外别作而来的。

4．证明中我们避免应用推论之二，因为从比例关系还须经过某些运算变化才能得出最后公式，而这种运算变化在我国古代文献中颇少依据，三上义夫在所著论《海岛算经》的一章中(见[18]，页 33～36)，曾经依据比例及其算法认为我国天元术代数学的产生，可从 13 世纪提前一千年至刘徽时期，这种过誉溢美

之辞，我们认为是不应接受的。

5. 我们的证明只需极简单的论证，从最基本的一些关系自然地直接获得刘徽原著所列举的那些形式上颇为复杂的公式。这足为这些证明基本上符合刘徽原意的一个旁证。与之相反，在利玛窦口译，徐光启笔受的《测量法义》中，以表测高的第10题实质上与海岛题相同，其证明颇有事先已知海岛公式而强欲用《几何原本》生搬硬套来获证，因而不无穿凿雕琢之嫌。由于《测量法义》一书不易见到，故将利玛窦的原图原证作为本文附录原样照录，以供大家参考。

以下将就《海岛算经》九题逐一证明。为了便利读者，原题以及术曰将依古原著照录，而在括号中注明附图相应的点或线段，诸公式即是原来术文的直接转述。

【一】望海岛

今有望海岛(AB)，立两表(DE、FG)齐高三丈，前后相去(EG)千步，令后表与前表参相直。从前表(DE)却行(EH)一百二十三步，人目著地(H)取望岛峰(A)，与表末(D)参合。从后表(FG)却行(GI)一百二十七步，人目著地(I)取望岛峰(A)，亦与表末(F)参合。问岛高(AB)及去表(BE)各几何?

术曰：

以表高乘表间(EG)为实，相多($GI-EH$)为法，除之。所得加表高，即得岛高。求前表去岛远近者，以前表却行乘表间为实，相多为法，除之，得岛去表里数。

望海岛公式：

$$岛高=\frac{表间\times表高}{相多}+表高,$$

$$(前)表去岛=\frac{表间\times前表却行}{相多}。$$

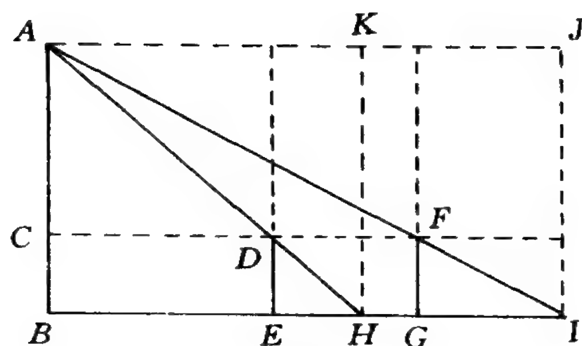


图 4

证 如图 4，从▭AI 与 F，得

$$▭FJ = ▭FB,$$

又从▭AH 与 D，得

$$▭DK = ▭DB。$$

相减，得 $$▭FJ - ▭DK = ▭EF,$$

或

后表却行×（岛高－表高）－前表却行×（岛高－表高）

＝表间×表高，

由此即得岛高公式。

又从▭DB＝▭DK 得

前表去岛×表高＝前表却行×（岛高－表高）。

应用岛高公式即得表去岛公式。

【二】望松

今有望松（AB）生山（BC）上，不知高下。立两表（DE、FG），齐高二丈，前后相去（EG）五十步，令后表（FG）与前表（DE）参相直。从前表却行（EH）七步四尺，薄地（H）遥望松末（A），与表端（D）参合。又望松本（B），入表（DJ）二尺八寸。复从后表却行（GI）八步五尺，薄地遥望松末（A），亦与表端（F）参合。问松高（AB）及山去表（CE）各几何？

术曰：

以入表乘表间(EG)为实，相多($GI-EH$)为法，除之。加入表，即得松高。求表去山远近(CE)者，置表间，以前表却行乘之为实，相多为法，除之，得山去表。

望松公式：

$$松高=\frac{表间\times入表}{相多}+入表,$$

$$表去山=\frac{表间\times前表却行}{相多}。$$

证　先将松高图(图 5)与海岛图相比较，视 AC 为海岛，则表去岛即此处之表去山 CE，故由海岛公式，得

$$表去山=\frac{表间\times前表却行}{相多}。$$

其次，将前表两测从松高图中割裂而作松高辅助图(图 6)，则从▭AH 与 D，得

$$▭DK=▭DC。$$

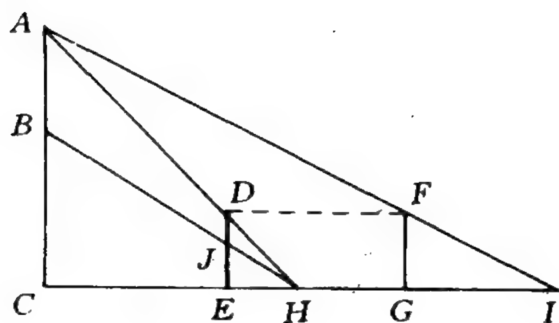


图 5

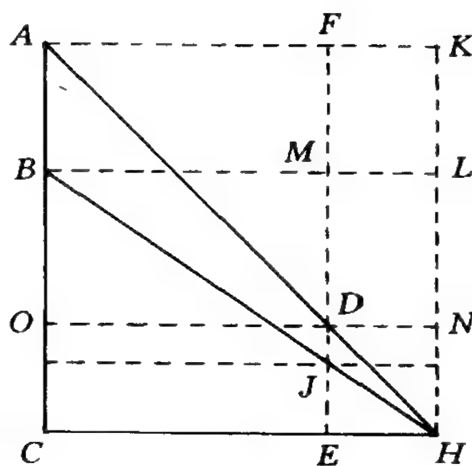


图 6

又从▭BH 与 J，得

$$▭JL=▭JC。$$

两式相减，得

$$▭MK-▭JN=▭JO。$$

或即

$$前表却行 \times 松高 - 前表却行 \times 入表 = 表去山 \times 入表。$$

由此得

$$松高 = \frac{表去山 \times 入表}{前表却行} + 入表。$$

这一公式在以后经常用到，将称之为松高辅助公式。以表去山公式代入这一辅助公式，即得所求松高公式。

【三】望邑

今有南望方邑（$ABCD$），不知大小。立两表（E，F）东西相去（EF）六丈，齐人目，以索连之。令东表（F）与邑东南隅（C）及东北隅（B）参相直。当东表之北却行（FG）五步，遥望邑西北隅（A），入索东端（IF）二丈二尺六寸半。又却北行去表（FH）十三步二尺，遥望邑西北隅（A）适与西表（E）相参合。问邑方（AB）及邑去表（BF）各几何？

术曰：

以入索乘后去表，以两表相去除之，所得为景差。以前去表减之不尽，以为法。置后去表，以前去表减之，余以乘入索为实，实如法而一，得邑方。

求去表远近者，置后去表，以景差减之，余以乘前去表为实。实如法而一，得邑去表。

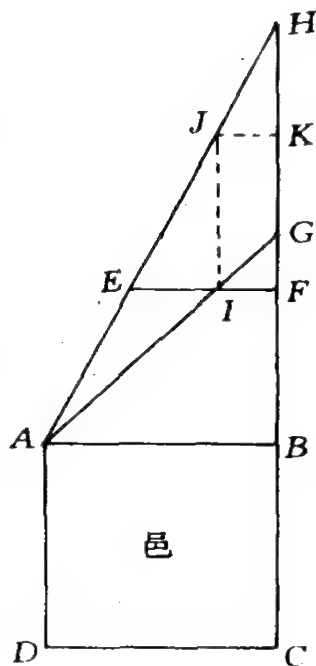


图 7

望邑公式：

$$景差 = \frac{入索 \times 后去表}{表相去},$$

$$邑方 = \frac{入索 \times (后去表 - 前去表)}{景差 - 前去表},$$

$$\text{邑去表} = \frac{\text{前去表} \times (\text{后去表} - \text{景差})}{\text{景差} - \text{前去表}}。$$

证 添作 IJ 与 JK，如图 7(望邑图)。从句股相似形 JKH 与 HEF，得

$$HK = \frac{JK \times FH}{EF} = \frac{IF \times FH}{EF}$$

$$= \frac{\text{入索} \times \text{后去表}}{\text{表相去}}。$$

称 HK 为景差，即得所求景差公式。

今将图 7 与望海岛图比较，视 AB 为海岛，IF, JK 为前后两表，则两图间的关系对照如下表所示：

	望　　邑	望海岛
AB	邑方	岛高
IF	入索	表高
$FK = FH - KH$	后去表 - 景差	表间
BF	邑去表	(前)表去岛
FG	前去表	前表却行
KH	景差	后表却行
$KH - FG$	景差 - 前去表	相多
$GH = FH - FG = FK + (KH - FG)$	后去表 - 前去表	表间 + 相多

依据这一对照表，岛高与表去岛公式，即各变为邑方与邑去表公式。

【四】望谷

今有望深谷(EFG)，偃矩(BAC)岸上，令句(BA)高六尺。从句端(B)望谷底(G)入下股(AD)九尺一寸。又设重矩($B'A'C$)于上，其矩间(AA')相去三丈。更从句端(B')望谷底(G)，

入上股($A'D'$)八尺五寸。问谷深(AE)几何?

术曰:

置矩间,以上股乘之,为实。上、下股相减,余为法,除之。所得以句高减之,即得谷深。

望谷公式:

$$谷深=\frac{矩间\times上股}{下股-上股}-句高。$$

证 如图 8(望谷图),从$\square B'G$ 与 D',得

$$\square D'H'=\square D'E。$$

从$\square BG$ 与 D,得

$$\square DH=\square DE。$$

相减得$\square D'I=\square D'A-\square DF$。

因 $\square D'I=\square A'I-\square B'D'$

$=\square BD-\square B'D'$

$=句高\times(下股-上股)$,

$\square D'A=矩间\times上股$,

$\square DF=\square DE-\square AF$

$=谷深\times(下股-上股)$,

故得 $谷深\times(下股-上股)=矩间\times上股-句高\times(下股-上股)$。由此得望谷公式。

图 8

【五】望楼

今有登山(AI)望楼(HG),楼在平地。偃矩(BAC)山上,令句(BA)高六尺,从句端(B)斜望楼足(G),入下股(AD)一丈二尺。又设重矩($B'A'C'$)于上,令其间相去($A'A$)三丈。更从句端(B')斜望楼足(G),入上股($A'D'$)一丈一尺四寸。又立小表($E'D'$)于入股之会(D'),复从句端(B')斜望楼岑端(H),入小

表（$F'D'$）八寸。问楼高（HG）几何？

术曰：上下股相减，余为法。置矩间，以下股乘之，如句高而一。所得，以入小表乘之，为实。实如法而一，即是楼高。

望楼公式：

$$楼高=\frac{矩间\times下股}{句高}\times入小表/(下股-上股)。$$

证 如图9（望楼图）。视GH为山HK上之松，$D'J$为前表，却行至B'以望松。则从望松辅助公式，得

楼高

$$=\frac{(山去楼-上股)\times入小表}{上股}+入小表$$

$$=\frac{山去楼\times入小表}{上股}。$$

从句股相似形$B'GI$与$B'D'A'$，得

山去楼/上股=（山高+矩间+句高）/句高。

视GI为深谷，则从望谷公式，得

山高+句高=矩间×上股/（下股-上股）。

由此诸式即得望楼公式。

图9

【六】望波口

今有东南望波口（AB）。立两表南（D）、北（C）相去（CD）九丈，以索薄地连之。当北表之西却行去表（CE）六丈，薄地遥望波口南岸（B），入索北端（CH）四丈二寸。以望北岸（A），入

前所望表里(GH)一丈二尺。又却后行去表(CF)十三丈五尺，薄地遥望波口南岸(B)，与南表(D)参合。问波口广(AB)几何?

术曰:

以后去表乘入索，如表相去而一。所得，以前去表减之，余以为法。复以前去表减后去表，余以乘入所望表里为实。实如法而一，得波口广。

望波口公式:

$$\text{波口广}=\frac{\text{入所望表里}\times(\text{后去表}-\text{前去表})}{\dfrac{\text{入索}\times\text{后去表}}{\text{表相去}}-\text{前去表}}。$$

证 补作 JK 等线，如图 10(望波口图)，今与望松题 2 相比较，令图中 $AB=$ 松，$AI=$ 山，$CH=$ 前表，$JK=$ 后表，则 $HK=$ 表间，$FJ-CE=$ 相多。在望松与望波口之间有下表中的关系。

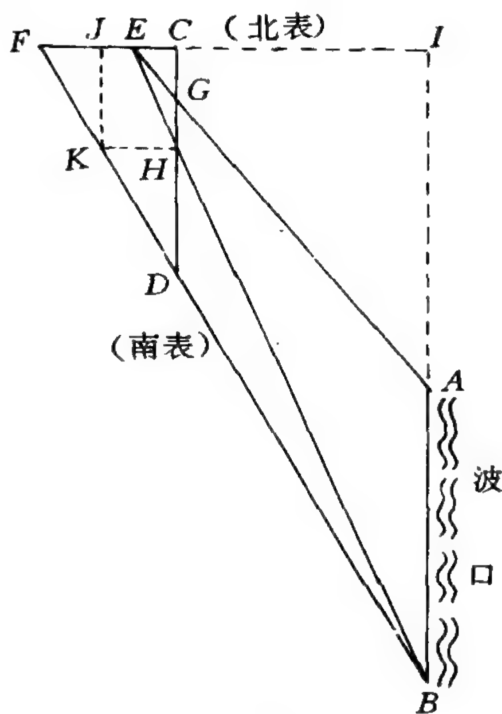


图 10

今依望松公式，有

$$\text{松高}=\frac{\text{入表}\times\text{表间}}{\text{相多}}+\text{入表}$$
$$=\frac{\text{入表}\times(\text{表间}+\text{相多})}{\text{相多}},$$

或即

波口广

$$=\frac{\text{入所望表里}\times(\text{后去表}-\text{前去表})}{\text{相多}}。$$

从句股相似形 FJK 与 FCD，有

$$FJ=\frac{JK\times CF}{CD},$$

或即

$$相多+前去表=\frac{入索\times 后去表}{表相去}。$$

'	望 波 口	望 松
AB	波口	松
$CH=JK$	入索	表高
GH	入所望表里	入表
EF	后去表－前去表	表间＋相多

以所得相多代入前波口广公式,即得所需公式。

【七】望清渊

今有望清渊,渊下有白石(G)。偃矩(BAC)岸(AH)上,令句(BA)高三尺,斜望水岸(F),入下股(AE)四尺五寸。望白石,入下股(AD)二尺四寸。又设重矩($B'A'C'$)于上,其间(AA')相去四尺。更从句端(B')斜望水岸(F),入上股($A'E'$)四尺。以望白石(G),入上股($A'D'$)二尺二寸。问水深几何?

术曰:

置望水上、下股,相减,余以乘石上股为上率。又以望石上、下股相减,余以乘望水上股为下率。两率相减,余以乘矩间为实。以二差相乘为法。实如法而一,得水深。

望渊公式:

$$水深=\frac{矩间\times(上率-下率)}{(望水下股-望水上股)\times(望石下股-望石上股)},$$

其中　上率＝望石上股×(望水下股－望水上股),

　　　下率＝望水上股×(望石下股－望石上股)。

证 如图 11(望渊图), FH 为水面, AH 为岸高, AI 为石深, FG 为水深。今视 GI 为深谷, G 为谷底, 则 AI 为谷深, 故依望谷公式, 有

$$石深 = \frac{矩间 \times 望石上股}{望石下股 - 望石上股} - 句高。$$

又视 FH 为深谷, F 为谷底。则 AH 为谷深, 故依望谷公式, 有

$$岸高 = \frac{矩间 \times 望水上股}{望水下股 - 望水上股} - 句高。$$

因

$$水深 = 石深 - 岸高,$$

故将上两式相减, 即得所求望渊公式。

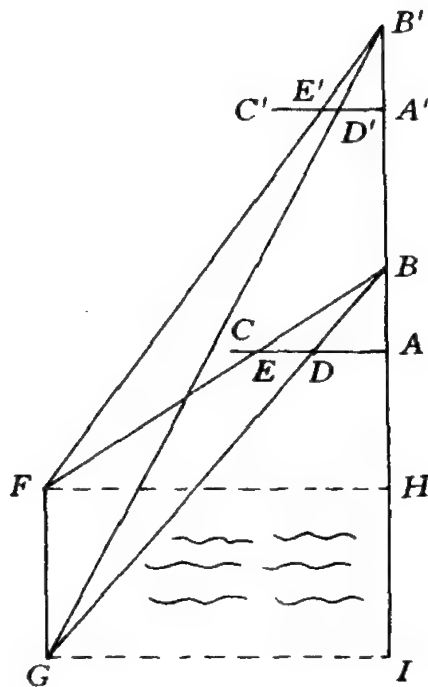


图 11

【八】望津

今有登山望津(FG), 津在山南。偃矩(BAC)山上, 令句高(BA)一丈二尺。从句端(B)斜望津南岸(F), 入下股(AD)二丈三尺一寸。又望津北岸(G), 入前望股里(DE)一丈八寸。更登高岩, 北却行(AH)二十二步, 上登($A'H$)五十一步。偃矩($B'A'C'$)山上, 更从句端(B')斜望津南岸(F), 入上股($A'D'$)二丈二尺。问津广(FG)几何?

术曰:

以句高乘下股, 如上股而一, 所得以句高减之, 余为法。置北行以句高乘之, 如上股而一, 所得以减上登。余以乘入股里为实。实如法而一, 即得津广。

望津公式：

$$津广=\frac{入股里\times\left(上登-\dfrac{北行\times句高}{上股}\right)}{\dfrac{下股\times句高}{上股}-句高}。$$

证　将图 12(望津图)与望松图相比较，视 FG 为松生山 GN 上，CA 与 KJ 为前后两表。由松高公式，得

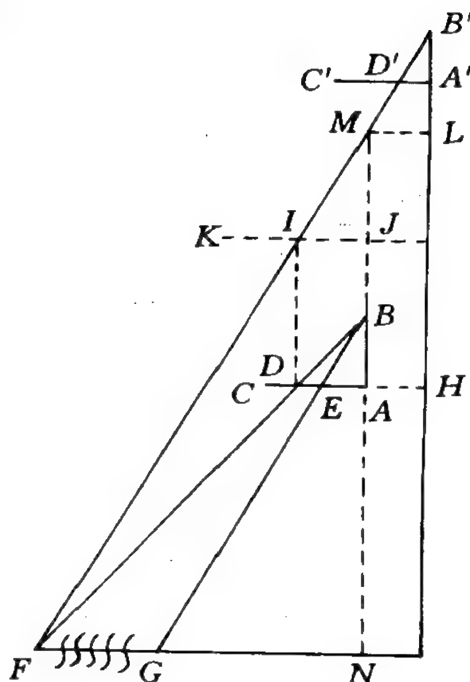


图 12

$$松高=\frac{入表\times表间}{相多}+入表$$

$$=\frac{入表\times(相多+表间)}{相多}。$$

式中松高 $=FG$ 即津广，入表 $=DE$ 即入股里，表间 $=AJ$。

又

$$相多=MJ-AB=MJ-句高。$$

$$相多+表间=MJ+AJ-句高$$
$$=LH-句高$$
$$=A'H-B'L$$
$$=上登-B'L。$$

故上式变为

$$津广=\frac{入股里\times(上登-B'L)}{MJ-句高}。$$

今从句股相似形 $B'LM$ 与 $B'A'D$，得

$$B'L=\frac{ML\times B'A'}{D'A'}=\frac{AH\times B'A'}{D'A'}=\frac{北行\times句高}{上股}。$$

又从句股相似形 MJI 与 $B'A'D'$，得

$$MJ=\frac{IJ\times B'A'}{A'D'}=\frac{AD\times B'A'}{A'D'}=\frac{下股\times句高}{上股}。$$

以之代入上津广公式，即得所求望津公式。

【九】 临邑

今有登山(AL)临邑($HIJK$)，邑在山南。偃勾(BAC)山上，令句高(BA)三尺五寸。令句端(B)与邑东南隅(J)及东北隅(K)参相直。从句端(B)遥望东北隅(K)，入下股(AE)一丈二尺。又施横句(EF)于入股之会(E)，从立句端(B)望西北隅(H)，入横句(EG)五尺。望东南隅(J)，入下股(AD)一丈八尺。又设重矩($B'A'C'$)于上，令矩间($A'A$)相去四丈。更从立句端(B')望东南隅(J)，入上股($A'D'$)一丈七尺五寸。问邑广(HK)、长(JK)各几何？

术曰：

以句高乘东南隅入下股如上股而一。所得，减句高，余为法。以东北隅下股减东南隅下股，余以乘矩间为实。实如法而一，得邑南北长也。求邑广，以入横句乘矩间为实。实如法而一，即得邑东西广。

临邑公式：

$$\text{邑南北长}=\frac{\text{矩间}\times(\text{东南下股}-\text{东北下股})}{\dfrac{\text{东南下股}\times\text{句高}}{\text{上股}}-\text{句高}},$$

$$\text{邑东西广}=\frac{\text{矩间}\times\text{入横句}}{\dfrac{\text{东南下股}\times\text{句高}}{\text{上股}}-\text{句高}}。$$

证 先将图 13(临邑图)与望松图相比较，视邑南北长 JK 为松生于山 KL 上，又视 DA 为表，从 B 点遥望，则

$$\text{入表}=DE=AD-AE=\text{东南下股}-\text{东北下股},$$

$$\text{表去山}=AL,$$

$$\text{前表却行}=AB=\text{句高}。$$

依松高辅助公式，有

$$松高=\frac{表去山\times入表}{前表却行}+入表=\frac{(表去山+前表却行)\times入表}{前表却行}$$

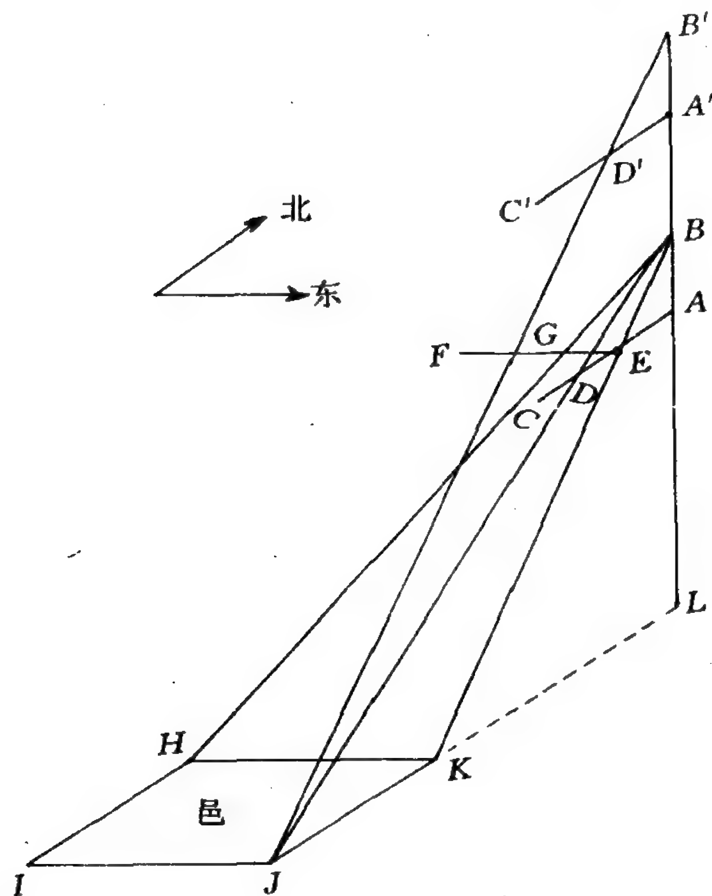


图 13

或

$$邑南北长=\frac{BL\times(东南下股-东北下股)}{句高}。$$

次将图 13 与望谷图相比较，视 JL 为谷，AL 为岸。则谷深 $=AL$，句高 $=AB$。故依望谷公式，有

$$BL=谷深+句高=\frac{矩间\times上股}{东南下股-上股}。$$

以此代入前式，即得所求邑长公式。

其次，在句股相似形 BHK 与 BGE 中，

$HK=$ 邑东西广，$GE=$ 入横句，故有

$$邑东西广 = \frac{BK \times 入横句}{BE}。$$

又由句股相似形 BKL 与 BEA，得

$$\frac{BK}{BE} = \frac{BL}{BA} = \frac{BL}{句高},$$

故得

$$邑东西广 = \frac{BL \times 入横句}{句高}。$$

将 BL 依前式代入，即得所求邑广公式。

附　　录

利玛窦口译徐光启笔受的《测量法义》第 10 题以表测高以及徐光启《测量异同》第四题以重表兼测无远之高无高之远，实质上即《海岛算经》的望海岛第一题，只是添上了人高而已，图 14 甲乙为欲测之高，癸壬、丁丙为表，丑子、己戊为人，利玛窦原证照录如下

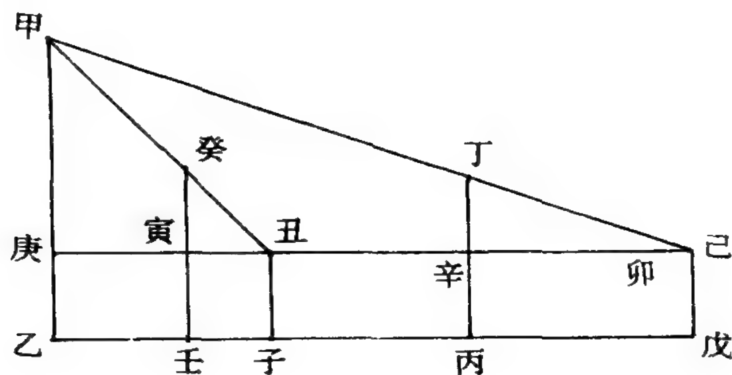


图 14

论曰，两测较卯辛与表目较辛丁或癸寅，其比例若距较戊子或己丑与庚甲。何者。己辛与辛丁既若己庚与庚甲[五卷四]，更之，即己辛与己庚若辛丁与庚甲也[五卷十一]，依显丑寅与丑

庚若寅癸与庚甲也，则丑寅与丑庚亦若辛丁与庚甲也，[辛丁与寅癸等故]，而己辛全线与己庚全线若己辛所截取之己卯，[己卯与丑寅等故]，与己庚所截取之丑庚也，则己辛全与己庚全亦若己辛分余之卯辛与己庚分余之己丑也[五卷十九]，前已论己辛与己庚若辛丁与庚甲即卯辛与己丑亦若辛丁与庚甲也。更之，即两测较卯辛与表目较辛丁若距较等子戊之己丑与甲庚也。若却后而得壬子，则反上论之。

后　　记

本文写成后重读李继闵同志的《从句股比率论到重差术》，李文提出对重差一词的解释以及我国古代测望理论“出入相补”──→“相似句股理论”──→重差术的发展过程，论证令人信服。但若仅用比例重差来证明，则海岛图不应再添许多线段以补成若干长方形，这显与杨辉、李潢所见的旧海岛图不同，一个可能是杨李所见的旧图并非刘徽所用的重差图。可惜刘徽的《重差图说》已经失传，难以印证。不论如何，这些证明在细节上略有出入，基本上都体现了我国古代几何学的固有特色，与以平行线等作为中心概念的欧几里得几何体系异其情趣，这应该是肯定的。

参考文献

[1] 钱宝琮校点，《算经十书》，中华书局，1963。
[2]《周髀算经》。
[3] 魏刘徽，《九章算术注》，263。
[4] 魏刘徽，《海岛算经》，263。
[5] 赵爽，《日高图说》。
[6] 宋杨辉，《续古摘奇算法》，1275。

[7] 利玛窦口译，徐光启笔受，《测量法义》，1607—1608。

[8] 明徐光启，《测量异同》。

[9] 清李潢，《海岛算经细草图说》，1820。

[10] 清李潢，《重差图序》，1820。

[11] 李俨，《重差术源流及其新注》《学艺》，1926，7卷8期，1—15。

[12] 李俨，《中国数学大纲》，科学出版社，1958，编Ⅰ章9。

[13] 钱宝琮，《中国数学史》1964。

[14] 白尚恕，《刘徽海岛算经造术之探讨》。

[15] 李继闵，《从句股比率论到重差术》，全国自然科学史学术会议，1980。

[16] Y. Mikami, The development of mathematics in China and Japan, 1913.

出入相补原理*

我国古代几何学不仅有悠久的历史，丰富的内容，重大的成就，而且有一个具有我国自己的独特风格的体系，和西方的欧几里得体系不同。这一几何体系的全貌还有待于发掘清理，本文仅就出入相补原理这一局部方面，就所知提出几点，主要根据是流传至今的以下各经典著作：

《周髀算经》（简称《周髀》），

《九章算术》（简称《九章》），

刘徽《九章算术注》（简称《刘注》），

《海岛算经》（简称《海岛》），

赵爽《日高图说》和《句股圆方图说》（简称《日高说》和《句股说》）。

田亩丈量和天文观测是我国几何学的主要起源，这和外国没有什么不同，二者导致面积问题和勾股测量问题。稍后的计算容器容积、土建工程又导致体积问题。我国古代几何学的特色之一是，依据这些方面的经验成果，总结提高成一个简单明白的一般原理——出入相补原理，并且把它应用到形形色色多种多样的不同问题上去。以下将列举这些不同的应用。

---

*本文摘自《中国古代科技成就》，中国青年出版社，1978年。

一、简单应用和比例理论

所谓出入相补原理，用现代语言来说，就是指这样的明显事实：一个平面图形从一处移置他处，面积不变。又若把图形分割成若干块，那么各部分面积的和等于原来图形的面积，因而图形移置前后诸面积间的和、差有简单的相等关系。立体的情形也是这样。

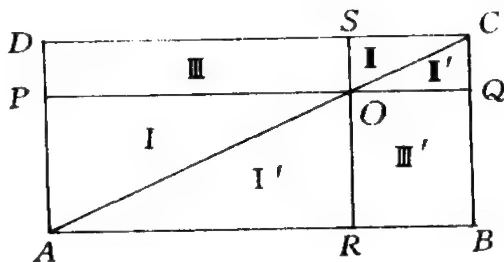


图 1

应用这一原理，容易得出三角形面积等于高底相乘积的一半这一通常公式，由此以定任意多角形的面积。作为另一简单实例，可以观察图 1，如果看作把$\triangle ACD$移置$\triangle ACB$处，又把Ⅰ、Ⅱ各移置到Ⅰ′，Ⅱ′，那末依出入相补原理，有

$$\text{Ⅲ} = \text{Ⅲ}', \ \square PC = \square RC。\cdots\cdots（指面积相等）$$

由此得

$$OP \times OS = OR \times OQ, \ PQ \times QC = RB \times BC, \ \cdots\cdots$$

因而 $AR : OQ = OR : CQ, \ AB : OQ = BC : QC, \ \cdots\cdots$

就是相似勾股形 ARO 和 OQC、ABC 和 OQC 的相应勾股成比例。并且可以导出其他相应部分的比例关系。

以上这些极简单的结果虽然没有在《九章》中明确说出，但是《九章》中曾经多处用这些关系来解决各种具体问题，参看《刘注》。

二、测望术和重差理论

在《周髀》中，就有用两表测日影以求日高的方法，公式是

$$日高=\frac{表高\times 表距}{影差}+表高。$$

如图 2，其中 A 是日，BI 是地平面，ED、GF 是先后两表，DH 和 FI 是日影。《海岛》改测日的高为测海岛的高，同图 AB 是海岛，H、I 是人目望岛顶和两表上端相参合的地方，于是日高公式成为：

$$岛高=\frac{表高\times 表距}{表目距的差}+表高。$$

刘徽证明和所用的图都已经失传，但是据现存《日高说》和残图以及其他佐证，原证当大致如下：

由出入相补原理，得

$$\square JG=\square GB, \tag{1}$$

$$\square KE=\square EB, \tag{2}$$

相减得 $\square JG-\square KE=\square GD$，

所以 $(FI-DH)\times AC=ED\times DF$，

即 表目距的差×（岛高－表高）＝表高×表距。

这就得到上述公式

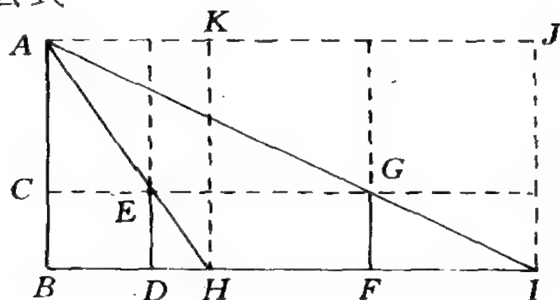


图 2

按《海岛》共九题都属测望之类，所得公式分母上都有两测的差，重差这一名称可能由此而来，其余八题公式都可依出入相补原理用和上面类似的方法证明，从略。

元朱世杰《四元玉鉴》中有和《海岛》完全类似的几个题，朱世杰对这些题的解法应该有古代相传下来的一定来历。依照朱对海岛一题的解法，我们认为原证比上面所示的可能稍复杂一些。如图 3，现在重作证明如下：

由出入相补原理，除（1）、（2）外，又有

$$\square PG = \square GD, \tag{3}$$

由（1）、（2）、（3），得

$$\square JN = \square EB = \square KE,$$

所以

$$IM = DH, \tag{4}$$

$$FM = FI - IM = FI - DH = \text{表目距的差。}$$

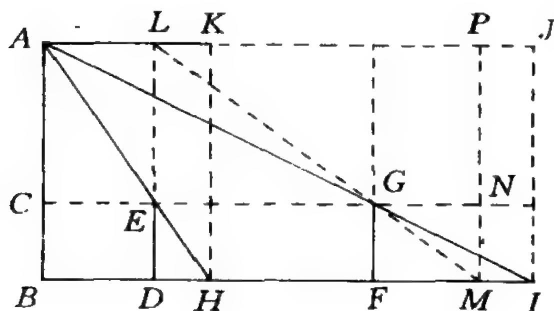


图 3

由（3）式就得到海岛公式。

如果依照欧几里得几何体系的习惯证法，那就自然应该添一平行线 $GM' \parallel AH$，如图 4，再利用相似三角形和比例理论作证。清代李潢以及近代中外数学史家大都依这法补作海岛公式证明，这当然不是刘徽的原意，也和我国古代几何传统相违背。注意作平行线的时候，应有 $FM' = DH$，和前面（4）式相比，M 和 M' 的位置完全不同。

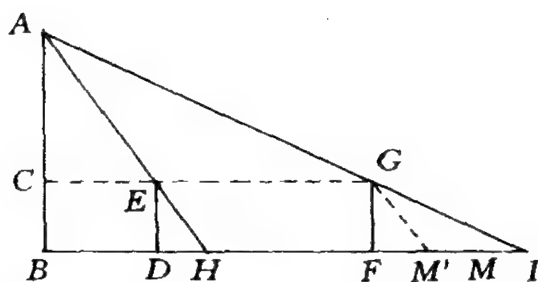


图 4

明末利玛窦来我国，他的主要学术工作之一是介绍欧几里得几何体系。他曾口授《测量法义》一书，其中载有和海岛题完全类似的一题。在他所作的证明中，需要在 FI 上取一点 M 使（4）式成立，再用比例理论作证，如图 4。按常理来说，利玛窦应该作平行线而取 M'，使得 $FM' = DH$，但是他一反欧几里得惯例，而和我国古代传统不谋而合，颇使人迷惑不解。现在提出这一问题，希望大家共同探讨。

三、勾股定理

在《周髀》和《九章》中，都已经明确给出了勾股定理的一般形式：勾^2 + 股^2 = 弦^2。虽然原证不传，但是据《句股说》以及《刘注》，都依出入相补原理证明，并且有遗留到现在可以用来作证的赵爽残图，这几方面互相参照，原证应该大致如下：如图 5，勾股形是 ABC，$BCDE$ 是勾方，$EFGH$ 是股方，把二者的和 $DB\ CF\ GH$ 的$\triangle BDI$ 移到$\triangle ABC$，$\triangle GHI$ 移到$\triangle AFG$，就得到 $ABIG$ = 弦方^2，由此得到勾股定理。

欧几里得《几何原本》中勾股定理的证明如图 6 所示，其中要先证有关三角形全等形以及三角形面积的一些定理，为此要做不少准备工作，因而在《几何原本》中直到卷一之末出现这一定理，而在整个《几何原本》中几乎没有用到。而在我

国，勾股定理在《九章》中已经有多种多样的应用，成为两千来年数学发展的一个重要出发点，参阅以下各节和文末附表。在东西方的古代几何体系中，勾股定理所占的地位是颇不相同的。

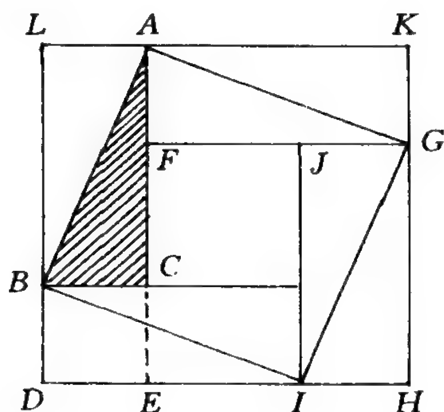
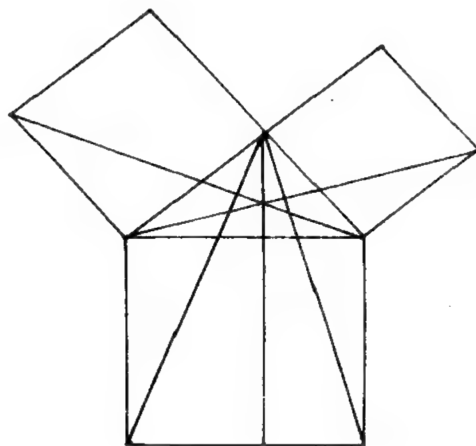


图 5


图 6

四、勾、股、弦及其和差互求

勾、股、弦和它们之间的和差共九个数，只须知道其中的两个就可以求得其他几个。除勾、股、弦互求就是开平方之外，《九章》勾股章中有不少这方面的问题。

1．知股弦差、勾，求股、弦（五题）。

2．知勾股差、弦，求勾、股（一题）。

3．知股弦差、勾弦差，求勾、股，弦（一题）。

4．知股弦和、勾，求股、弦（一题）。

各题都列出了一般公式，《句股说》的许多命题也属这一类，《刘注》还给出了证明，公式的来历和证明的方法都依据出入相补原理，有的也用比例原理作别证。试以勾股章第十三折竹题为例。题设竹高已知，竹在某处折断，竹梢着地，着地处和竹根距离也已知。求折断处的高度，如图 7。如果以着地处和

竹根的距离作为勾，就是从股弦和、勾求股的问题，《九章》原文给出的公式是

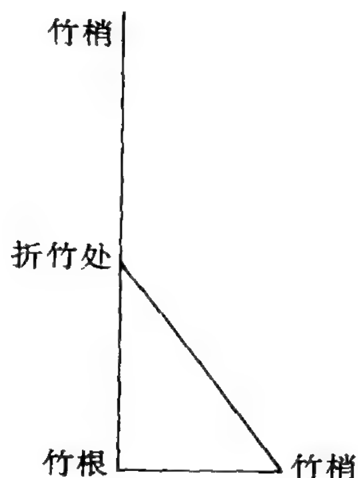


图 7

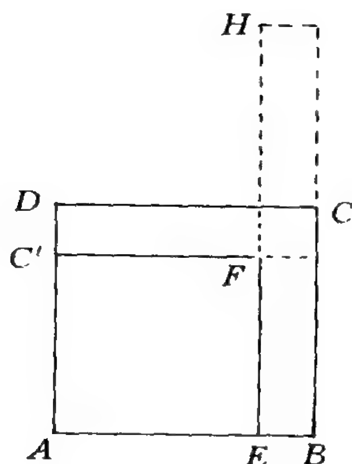


图 8

$$股弦差 = 勾^2/股弦和,$$

《刘注》又给出了另一公式：

$$股 = \frac{股弦和^2 - 勾^2}{2 \times 股弦和}。$$

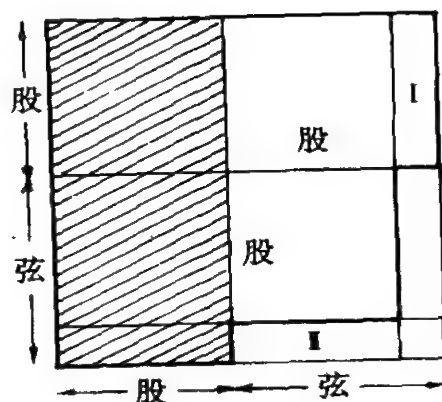


图 9

为了证明前一公式，可以考虑图 8，其中正方形 $ABCD$ 和 $AEFC'$ 的边各是勾股形的弦和股。依勾股定理曲尺形 $EBCDC'F$ 的面积应该等于勾^2。现在把▭FD 如图移到▭CH，那末依出入相补原理，▭BH 的面积是勾^2，而它的边长各是股弦和、股弦差，就得到上面的前一公式。

另一公式的刘徽证明也相类似。试考察图 9，其中右下角曲尺部分面积依勾股定理等于勾^2，所以粗黑线围成部分等于股弦和^2 - 勾^2。把长方形Ⅰ移到Ⅱ，依出入相补原理，这一面积是斜线部分的两倍，就是 2×股×股弦和，由此就得到另一

公式。

五、秦九韶公式

秦九韶《数书九章》中有一题是已知不等边三角形田地三边的长（称大斜、中斜、小斜，以下简记为大、中、小），求田地面积。秦九韶的解法相当于下面的一般公式：

$$面积^2=\frac{1}{4}\left[小^2\cdot大^2-\left(\frac{大^2+小^2-中^2}{2}\right)^2\right]。$$

秦的公式来历不明，证明也失传了。

现在补作一证如下：

作大斜上的高分大斜成两部分，作为勾股形的股和弦，如图10。由于《九章》已给出三角形面积等于 $1/2\times高\times大$，所以问题归结为怎样求高，或怎样求股。由于

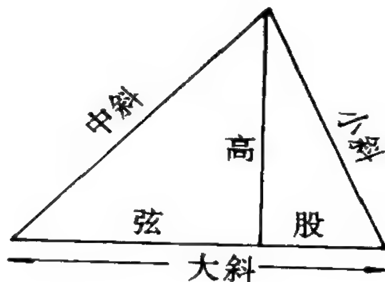


图 10

$$股弦和=大,$$

$$勾^2=弦^2-股^2=中^2-小^2,$$

所以问题归结为怎样从股弦和、勾求股。依上节的刘徽公式，得

$$股=\frac{股弦和^2-勾^2}{2\times股弦和}=\frac{大^2-(中^2-小^2)}{2\times大},$$

$$高^2=小^2-股^2=小^2-\left(\frac{大^2+小^2-中^2}{2\times大}\right)^2。$$

由此就得到秦的公式。

按秦公式的形式十分古怪，当是依某种思路自然引导到这一形式的。上面的证法颇为自然，也符合我国古代几何的传统特色，说它是原证，也是不无可能的。

在西方有所谓海伦公式（a、b、c 是三角形三边的长）：三角形面积 $=\frac{1}{4}\sqrt{(a+b+c)(b+c-a)(c+a-b)(a+b-c)}$，这一公式形式十分漂亮。正因为这样，如果已知海伦公式而再来推出秦的公式，将是不可思议的，相反，从秦的公式化简成海伦的公式，乃是比较自然的发展。

据此我们至少可以断言，秦的公式是独立于海伦公式而得来的。

关于海伦的生平，从公元前 2 世纪到公元后 10 世纪以后，数学史家聚讼纷纭。至于海伦留传到现在的著作，历代都重新编纂过，有所增改，已经不是本来面目。这是熟悉希腊数学史的应予澄清的事，这里就不考虑了。

六、开平、立方

从勾、股求弦，先把勾、股平方后相加，再开平方就得弦。因而勾股定理的应用自然导致开平方的问题。事实上，《周髀》中已经给出了若干具体数目的平方根，而在《九章》中，更详细说明了开平方的具体方法和步骤。这一方法的根据是几何的，就是出入相补原理。试以求 55225 的平方根为例。这相当于已知正方形 $ABCD$ 的面积是 55225，求边 AB 的长，如图 11。按我国记数用十进位位值制，因 AB 显然是一个百位数，所以求 AB 的方法就是依次求出百位数字、十位数字和个位数字。先估计（《九章》中用“议”字）百位数字用 2，因而在 AB 上截取 $AE=200$，并且作正方形 $AEFG$，它的边 EF 的两倍称为“定法”。把 $AEFG$ 从 $ABCD$ 中除去，所余曲尺形 $EBCDGF$ 的面积是 $55225-200^2=15225$。其次估计十位数字是 3，在 EB 上截取 $EH=30$，并且补成正方形 $AHIJ$。从

$AEFG$ 所增加的曲尺形 $EHIJGF$ 可以分解成三部分：▭FH，▭FJ，▭FI，面积依次是 $30\times EF$，$30\times FG$，30^2，其中 $EF = FG = 200$，所以从 $ABCD$ 中除去 $AHIJ$，所余曲尺形 $HBCDJI$ 的面积是

$$15225-(2\times 30\times 200+30^2)=2325。$$

现在再估计个位数字是 5，在 HB 上截取 $HK=5$，并补作正方形 $AKLM$，从 $ABCD$ 中除去后所余曲尺形面积，和前同法应该是

$$2325-(2\times 5\times 230+5^2)=0。$$

由此知 K 和 B 重合，而 55225 的平方根恰好是 235。

求立方根的方法和步骤和这相似，但是要把一立方体逐步进行分解，比平方根求法稍复杂，所依据的仍是出入相补原理。这在《九章》中也有详细叙述。

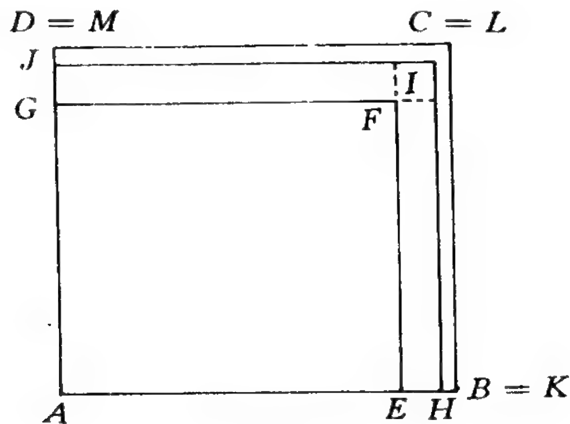


图 11

我国开平立方法来源很古，它的几何本质十分清晰，而且从方法上可以看出我国独有而世界古代其他民族所无的位值制记数法的高度优越性。不仅这样，至迟到 11 世纪中叶，我国就已经把开平立方法推广到开任何高次幂，就是所谓“增乘开方法”，并且出现了有关的二项式定理系数表，就是所谓“开方作法本源图”。从这一方法的几何渊源看来，如果说当时的

我国数学家已经有高维方体和高维几何的雏影，似乎不是全无根据的。

七、解二次方程

在开平方的过程中，曾经出现象图 8 中黑线部分那样的图形，其中 $2\times EF$ 称定法。开平方在求得 AE 后，其次几步在于从曲尺形 $EBCDC'F$ 的已知面积求得 EB。现在把▭DF 移到▭CH，那末依出入相补原理，▭BH 面积已知，此外▭BH 的两边 EH 和 EB 的差就是定法 $2\times EF$，也有已知数值。因而求 EB 的问题可以转化为下面的问题：

(A) 已知一长方形（▭BH）的面积、长阔差，求长阔。

反过来，这一问题的解法，可依开平方中第二步以下的方法求得，称为“开带从平方”。自《九章》以来用下面的语句来表达：

(B) 以（长方形面积）为实，（长阔差）为从法，开方除之，得（阔)”。

以上“从法”一名，当来自开平方过程中的“定法”，“开方”一词也说明了它的来历。

下面的例取自《九章》，如图 12。图中 $ABCD$ 是一方城，出北门北行若干步到 G 有木，出南门南行若干步到 F 再西行若干步到 H，恰可望见木 G，问题是求方城每边的长。据《刘注》的方法是依出入相补原理得▭$EJ=2$▭$EG=2$▭$KG=$ $2\times$北步$\times$西步。又▭EJ 的长阔差是南步＋北步。所以解法是以 $2\times$北步$\times$西步为实，以南步＋北步为从法，开平方除之，得 EI，也就是方城边长。

不仅应用开平方法可得问题（A）的数值解，而且应用出入相补原理，还可以求得解答的精确表达式。如果以长方形的

阔为勾，长为股，那末问题（A）相当于：

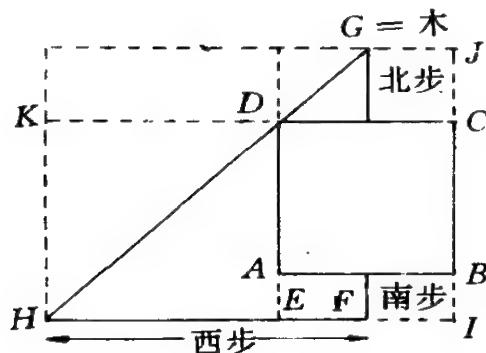


图 12

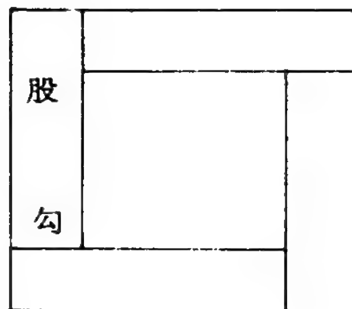


图 13

（C）已知勾股积、勾股差，求勾、股。

为此考虑赵爽残图如附图 13。图中大小两正方形的边长各是勾股和、勾股差，所以得

$$勾股和^2 = 4\ 勾股积 + 勾股差^2。$$

由此得勾股和，因而得勾和股。同样也可从勾股和、勾股积求得勾和股，这一方法可以参阅《勾股说》的末一命题。

宋元时期明确引入了未知数的概念。如果以 x（当时称为天元一）表长方形阔，那末问题（A）相当于解一个二次方程

$$x^2 + ax = b,$$

其中 a 相当于从法，b 相当于实。所以在古代实质上已经给出了这一形式二次方程（a、b 都是正数）的近似解和精确解，前者在宋元时期发展为求任意高次方程的数值解法，后者虽文献散佚不可查考，但是据唐初王孝通的著作以及史书关于祖冲之引述看来，不能排除我国曾经对三次方程用几何方法求得精确表达式的可能性。参阅钱宝琮《中国数学史》关于祖冲之开差幂、开差立的讨论。

在其它各国，9 世纪花刺子摸的代数学名著中列举了各种类型二次方程的精确解法，它的方法是几何的，它的精神实质和出入相补原理颇相类似。16 世纪意大利数学家关于三次方

程解法，也完全是几何的。

八、体积理论和刘徽原理

如果规定长方形的面积是长阔的积，那末依据出入相补原理，容易得到：

（1）三角形面积 $=\frac{1}{2}\times$ 高 $\times$ 底。

由此可以完全奠定平面多角形的面积理论。但是在空间情形，如果规定长方体的体积是长、广、深的积，是否依据出入相补原理，可以推得

（2）四面体体积 $=\frac{1}{3}\times$ 高 $\times$ 底面面积，

由此以建立多面体的体积理论，就不是那末明显而是极为困难的问题。欧洲直到 19 世纪末，才把它作为一个难题明确地提出来。1900 年希尔伯特在国际数学会上所作著名讲演中，把体积理论列为 23 个问题之一。这一问题立即为德恩所解决，答案是否定的：两个多面体要分割成彼此重合的若干多面体，必须满足某些条件，通称德恩条件。自此以后直到 1965 年，一位瑞士数学家西德勒才证明了德恩条件也是充分的。但是问题决不能认为已经彻底解决。从希尔伯特直到晚近，多面体体积理论仍不断成为一些知名数学家讨论的课题。德恩条件叙述复杂，也难以认为是合宜的最后形式。

在这种情势下，看看中国古代对这一问题的处理方式是不无有启发性的。

《九章》以至《刘注》解决体积问题的出发点是把一般的多面体分解为一些基本的立体。先把一长方体斜剖为二，如图 14（1），得两堑堵，如图 14（2），再把堵堑斜剖为二，一个是阳马，如图 15（1），一个是鳖臑，如图 15（2），其中鳖臑

的特征是 AB 和平面 BFG 垂直，FG 和平面 ABF 垂直。由于任一多面体可以分割为四面体，而任一四面体可以分割为 6 个鳖臑，如图 16，所以问题归结为求鳖臑（以及阳马）的体积。依刘徽原话，就是所谓阳马、鳖臑，“功实之主也”。

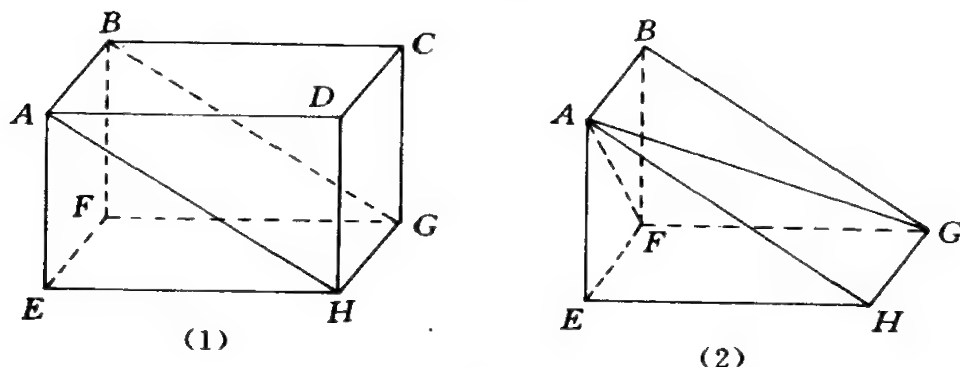


图 14

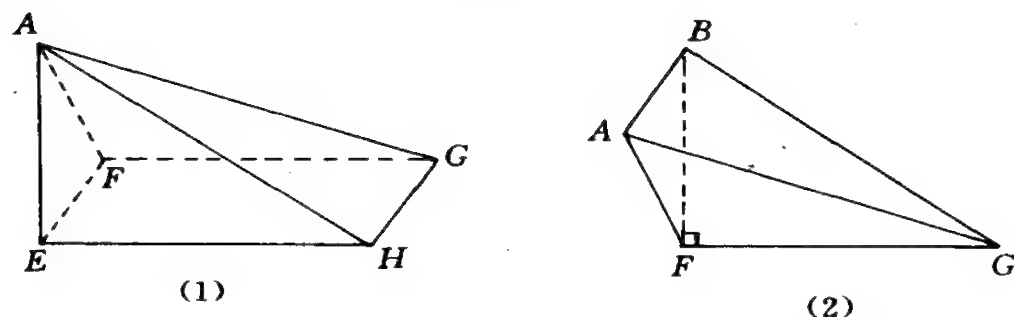


图 15

其次的问题是怎样求得阳马和鳖臑的体积。如果长方体成为立方体，那么分解所得的阳马的体积是鳖臑的两倍。刘徽作了长篇的分析，得出结论是：这个论断普遍成立，用刘的原话是：“阳马居二，鳖臑居一，不易之率也。”我们把它称为

刘徽原理 斜解一长方体，所得阳马和鳖臑的体积的比恒是二比一。

从这一原理容易得到鳖臑和阳马的体积公式。由此又容易得到（2）式，因而整个多面体的体积理论可奠基于刘徽以及出入相补这两个原理之上。

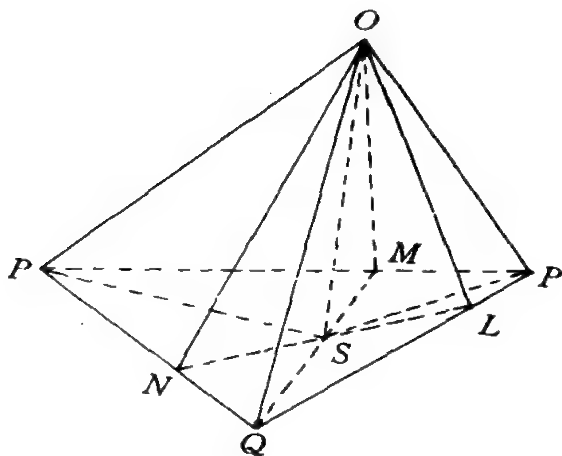


图 16

刘徽对自己原理有详细的分析说明，实际上就是这一原理的证明。遗憾的是因为年代久远，文字有错误脱落，已经难以索解，但是无疑是一极限过程的证法。把这一段彻底弄清，将是一件很有意义的工作*。按希尔伯特和他的后继者们的研究指出，体积理论和面积理论不同，除要有出入相补原理外，必须辅以连续一类公理。也有人（例如沙顿诺斯基，1903 年）提出排除连续公理，直接应用（2）式作为建立体积理论的基础。但是这样就要先证明（2）式中高和底面积的乘积凡四都彼此相等，这既不明显也不简单，似不如刘徽原理和出入相补原理的显豁自然。

总之，多面体的体积理论到现在还余蕴未尽，估计中国古代几何中的思想和方法，或许对进一步的探讨还不无帮助。

九、羡除公式

《九章》中列举了各种多面体的体积，依据的就是出入相

* 这一段事实上早已弄清楚，参阅近年来丹麦华道安与自然科学史研究所郭春书的一些论著以及有关论文。

补原理和阳马鳖臑公式。现在以羡除即隧道（图 17）为例，图中 $ABCD$ 是地面，成一梯形，$CDEF$ 是隧道的一端，成垂直平面中的梯形。整个隧道依剖面 IJK 对称。EG、FH 和 CD 垂直是隧道的深，IJ 是隧道地面的长，CD、EF、AB 各称上广、下广、末广。《九章》给出的公式是：

$$\text{羡除体积} = \frac{1}{6}\ (\text{上广} + \text{下广} + \text{末广})\ \times \text{深} \times \text{长}。$$

《刘注》的证法是先把羡除分解，如在图 17 中 $CD > AB > EF$ 的情形，分解成一个堑堵 $EFGHLM$，两个小鳖臑 $AGEL$ 和 $BFHM$，两个不正规大鳖臑 $ACEG$ 和 $BDFH$，再应用堑堵、鳖臑公式和上一节公式(2)，就得到这一公式。这一方法在《九章》中用来求得例如刍甍、刍童、盘池、冥谷等多面体的体积公式。

如果依 IJK 剖面取羡除的一半，所得 $IJKACE$ 如图 18 是

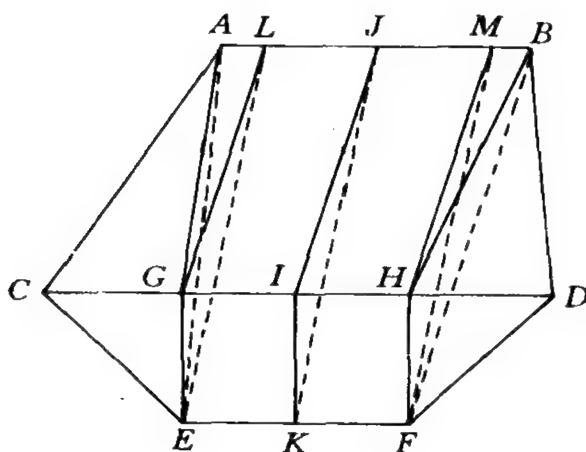


图 17

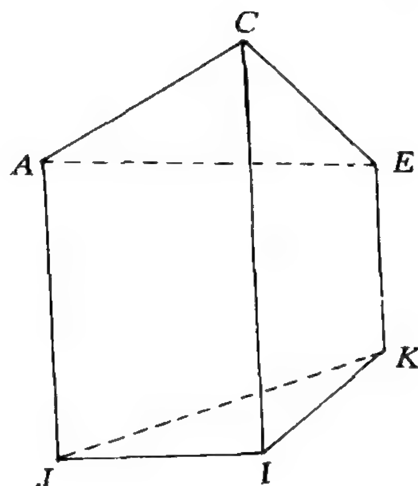


图 18

一斜截直柱体，是把一个以勾股形为底面的直柱体斜截而成，它的体积是三高平均值和底面面积的积。因由任意曲面所围成的立体可以看作近似地由这样的斜截直柱体构成，所以据此可以得出函数 $f(x, y)$ 的积分近似公式，犹之微积分中求曲线下面积的辛普森积分近似公式。因而羡除公式具有重要意义。

在西方，斜截直柱体的体积公式最早见于 1794 年勒根德尔所著《几何原理》一书，因此也称为勒根德尔公式。勒根德尔一书是从欧几里得《几何原本》以后最早可以代替《几何原本》的名著，它的有关公式的证明同样依据四面体体积公式，但是它的分解方法和《刘注》不同。此外某些多面体西方也有不同的分解法和证法，不妨中外参照，加以比较。

十、球体积和祖暅原理

从《九章》到《刘注》，我国对多面体的体积已经建立了相当完整的理论体系。但是对于求曲面围成的立体，特别是球的体积问题，却遇到了困难。这一球体积问题，直到南北朝祖暅才完全解决，为此并且提出了所谓

祖暅原理　幂势既同，则积不容异。

这一原理在公元 17 世纪以卡瓦利里原理的形式重见于欧洲，成为微积分得以创立的关键性的知识。

祖暅关于球体积公式的证明见于《九章》的唐李淳风注，论证极其详细清晰。证明分三步：

（1）在一立方体中依两不同方向作两内切圆柱体，它的共同部分称“牟合方盖”。取立方体的$\frac{1}{8}$，其中方盖部分称内棋，此外三部分称外棋。依祖暅原理可得

$$\frac{1}{8}\text{球体积}:\frac{1}{8}\text{方盖体积}=\pi:4。$$

（2）从立方体的$\frac{1}{8}$割出一倒立的阳马，应用勾股定理证得三外棋等高处截面积的和跟阳马同高处的截面积相等。

（3）再应用祖暅原理，知三外棋体积的和跟阳马体积相等。

由阳马的体积公式，就可以从上述三步得球体积公式。

按牟合方盖是刘徽所引入的，第一步的结果实质上也已经为刘徽所求得。事实上，在《刘注》中，他已经多次应用了祖暅原理来求曲面围成立体的体积，例如从方堡垛求圆堡垛，从方锥求圆锥，从方亭求圆亭，都已经使用这方法。祖暅的功绩，不仅在于具体求出了牟合方盖因而求出球的体积，更在于把实际上已知并且已经广泛应用的实践经验总结提高到一般原理的形式。是否应该把祖暅原理改称为刘祖原理，是可以商讨的。

从祖暅原理可以立即得出前面讲到的刘徽原理，因而多面体的体积理论也可以建立在出入相补原理和祖暅原理这两个浅显易明的基本原理之上。在欧洲，直到希尔伯特的《几何基础》问世以后，20 世纪初年，才有人（例如绪思）考虑依卡瓦利里原理以建立体积理论的问题。

十一、其　　他

《九章》中有丰富的几何学内容，即使局限于出入相补原理，除了已经见于前面的各节的以外，还有一些成果成为我国数学以后发展的重要出发点。例如所谓勾股容圆问题，在李冶的《测圆海镜》中已经有了很大的发展。又如前面提到过的所谓方城问题，在秦九韶、李冶等著作中已经把方城改成了圆城，这是旧有方法所不能解的。为此宋元时期创立了所谓天元术一类新的理论和方法，不仅可以用来解决许多新问题，对老的问题（所谓古问）也提供了新的有力工具，和老的方法（所谓古法）相比可以“省功数倍”。这些新理论新方法的实质在于几何的代数化，乃是解析几何的前奏，也是近代代数学的前驱。

出入相补、刘徽、祖暅等一般原理的建立，说明我国古代学者具有高度的抽象概括能力，善于在深入广泛的实践基础上

往高里提。这些原理之简单易明正可和它们的应用之广互相辉映。这是我国古代数学的一种独特风格，着重在问题的解决以及解决的一般方法和一般原理原则。同样的风格也可见之于几何的代数化、位值制记数法等等。这和西方数学之偏重于概念和概念之间的相互逻辑关系，是异其旨趣的。

我国数学经典著作散佚的多而保存的少，就象祖暅原理，也只靠李淳风一注才得以留传下来。象这一类重要成果而失传无从考查的，当不在少数。尽管如此，只从留传至今的典籍看来，我国数学的生产实践方面的渊源和发展演变的线索，仍旧很分明，参见下面两个附表。

表 一

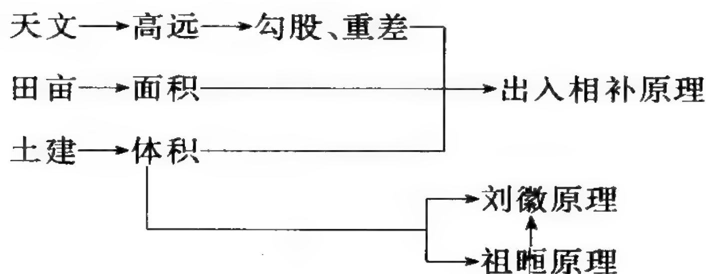
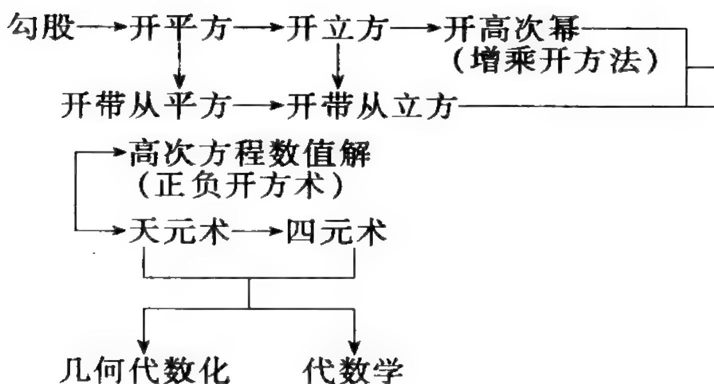


表 二



《〈九章算术〉注释》的序*

《九章算术》是我国古代流传下来的一部数学巨著，不仅指导着我国数学的发展达两千余年之久，而且对世界数学的发展也有不可估量的巨大影响，线性联立方程组的解法及有关正负数的引入祇其一例而已。我国古代数学有它自己的体系与形式，与西方之以欧几里得几何为代表的所谓公理化体系者旨趣既异，途径亦殊。《九章算术》与《几何原本》东西辉映，无疑是数学史上的两大传世名著，也是现代数学的两大源泉。

《九章算术》的刘徽注是数学上的又一伟大成就。刘徽注不仅提出了丰富多采的创见与发明，并以严密的数学用语描述了有关数学概念，对《九章算术》中的许多结论给出了严格证明。他所采用的证明方法，不仅有综合法、分析法，而且有时还兼用反证法。他沿袭我国古代的几何传统，使之趋于完备，形成具有独特风格的几何体系。刘徽的发明、创造对后世人有所启发，即使对于现今数学也有不少借鉴之处。从对数学贡献的角度来衡量，刘徽应该与欧几里得、阿基米德等相提并论。

遗憾的是，像传本《九章算术》与刘徽注这样的伟大著作，由于古今文字迥异，专门名词与现代通用者更大不相同，加上文字简略，用字深奥，使当代有志者难于领略。白尚恕同志博征详考，对全书用现代通俗易懂的语言详加注释，既使国内外对我国古代数学有兴趣的人士易于涉猎了解，也使研究我

---

* 本文摘自《〈九章算术〉注释》(白尚恕著)，科学出版社，1983年。

国古代数学的发展及刘徽与其他人如李淳风等的创见有途可循，为之称便。这是一件十分有意义的事。为此不揣冒昧，谨志数语，聊为此书出版作贺。

《〈九章算术〉与刘徽》序*

《九章算术》是我国数学方面流传至今最早也是最重要的一部经典著作。它承前启后，一方面总结了秦汉以前的数学成就，另方面又成为汉代以来达两千年之久数学研究与创造的源泉。特别是三国时期魏刘徽的《九章注》，对数学理论多所阐发，影响深远。总之，《九章算术》与刘徽《九章注》，对数学发展在历史上的崇高地位，足可与古希腊的《欧几里得几何原本》东西辉映，各具特色。本书汇集了 19 篇专题论文，从各个角度对这一部传世杰作进行研讨。在这些论文之前，冠以《出版小志》一篇，就《九章》的成书背景、内容版本、注释校证与对后世的影响，以及最重要的注释者刘徽的事迹作了概括性的介绍。

《九章算术》与刘徽《九章注》源远流深，不仅对我国古代数学的发展，即使对整个世界数学的发展也有巨大影响，《九章》第八方程章的线性联立方程组解法与正负数概念的引入，祇是比较显著的一例而已。要把《九章》在世界数学中的地位，与世界其它地域数学发展的关系及影响的来龙去脉弄清，还需做大量细致的研究调查工作。本书所收集的一些论文，还祇能算是一个开端，真正的艰巨工作还在今后。为了便利读者作进一步的深入研究，以及为了便利与国外的交流，由

* 本文摘自《〈九章算术〉与刘徽》(吴文俊主编)，北京师范大学出版社，1982年。

白尚恕、李迪、沈康身三位同志写成了名词今释、论文索引与英文提要三篇，添入附录。希望读者们能对本书多提意见，使今后的工作能继续向纵深开展。

《秦九韶与〈数书九章〉》序*

出现在秦汉之际的《九章算术》，是一部综合当时数学成就的经典巨著。自此以迄宋代，虽历代皆有著作，其中未遭散佚幸存至今者或则创见不多，或则仅及一隅（如《缉古算经》）。其能论述全面而富有创造性成就能与《九章算术》相比美者厥唯秦九韶《数书九章》一书。秦书之大衍求一术与增乘开方术，都已成为经典。夫"术"虽有时也代表公式或定理，但在大多数场合即犹今之"算法"。我国古算往往寓理于算，而以机械化的思想方法为其特色。求一算法与开方算法，即为体现这种思想方法的两大辉煌成就。这与西方数学之以演绎推理为主的公理化体系正相对照而互相辉映。秦汉之《九章》与宋代之《九章》，正是综述这种机械化思想体系所获成就的两大巨著。为复兴我国固有的传统数学，继承发扬其特色，以振兴我国未来的数学事业，有识之士正可从其中吸取力量。今继《〈九章算术〉与刘徽》一书之后，续以《秦九韶与〈数书九章〉》。仍按前者体例，汇集论文30篇，从各个不同角度论述秦九韶其人、其书，以志不忘古人创业之艰、贡献之巨，并以勉后来者。是为序。

---

* 本文摘自《秦九韶与〈数书九章〉》（吴文俊主编），北京师范大学出版社，1987年。

《郭书春汇校〈九章算术〉》序*

相传约在公元前三百年左右，地中海岸亚历山大城的欧几里得写成了《几何原本》（以下简称《原本》）一书，代表了古希腊数学的一个顶峰。至于该书所依据更古时期的文献，则除了一些不足道者外，已无整篇留传至今。学者们从现在各种资料中企图拼凑出欧几里得其人及其以前的数学面貌，大都有臆测成分。这与巴比伦泥版之类地下文物之确凿可据，是无法比拟的。

在中国，则史称隶首作数，至周公制礼而有九数，汉代郑玄为《周礼》作注，曾提及九数及若干种篇章的细目。秦汉之际张苍、耿寿昌因旧文删补，至公元前后定型而成《九章算术》（以下简称《九章》）一书，其内容与郑玄所述基本上无大差异，即刘徽所谓"九数之流，则《九章》是矣。"因而《九章》应可视为周公以至秦汉我国在数学上成就的一个大总结。近年来的地下发掘，所发现的大量文物，更为《九章》前身提供了实物佐证。

《原本》开创了公理化演绎系的纪元。其思想与方式，在数学的现代研究中占据着一种统治地位。但从《原本》出现以至今日，其经历不无坎坷曲折。在漫长的中世纪时期，《原本》以至整个希腊学久已默默无闻。直至9世纪时，才由阿拉伯人将《原本》的希腊文传成阿拉伯文，至12世纪的十字军战争

---

* 本文摘自《郭书春汇校〈九章算术〉》，辽宁教育出版社，1990年。

期间，又从阿拉伯文译成拉丁文。至于拉丁与希腊文本在欧洲的广泛流行，则似有待于 15 世纪以后。

在我国，则《九章》以其独特的方式与方法，阐扬了以算为主以术为法的算法体系的泛滥，《九章》之义更为一般数学家所不屑。但由于近代计算机的出现，其所需数学的方式方法，正与《九章》传统的算法体系若合符节。《九章》所蕴含的思想影响，必将日益显著，在下一世纪中凌驾于《原本》思想体系之上，不仅不无可能，甚至说是殆成定局，本人认为也决非过甚妄测之辞。

现在《原本》，决非欧几里得原来的写本。现在所见《原本》的各种抄本中，最早的希腊文抄本在公元后 10 世纪时，距今人近而距欧几里得远。其间已不知经历过多少遍辗转传抄。丹麦学者 Heiberg，曾以难以置信的毅力，对现存各种《原本》的版本以科学方法细加勘校，所得结论是：所有版本除一种外，都导源于公元后 4 世纪时埃及著名学者亚历山大城 Theon 的一个抄本。至于孤立的另一抄本，则其出现的时间应在 Theon 的抄本之后。由于欧洲习惯在传抄时往往以己意随意添加删补而不加注明，因而现存《原本》与欧几里得原来写本有多少出入，特别是 Theon 对原著的忠实程度如何，都是难以判定的。

我国的《九章》也同样历经传抄、注释与刊印，存在着多种不同的版本。与欧洲的习惯不同，中国的史家有着优良的传统。所有注释家如刘徽、李淳风等，都须严格注明何处是本人加注，何处是原文所有，经纬分明，不容混淆。因而传抄虽难免有误，原文实质可保证妄加纂改之虞，鉴于《九章》在数学发展历史上的已有作用以及对未来无可估量的影响，理应对各种版本细加校勘。对辗转传抄与刊印中可能出现的谬误一一指出，如 Heiberg 之于《原本》所为，应是一件不容回避的重要

工作。郭书春同志多年来博采群书，艰苦奋尝，终于完成了这一艰巨的历史性任务。在此书行将出版之际，特书此以聊表庆贺之情。

《〈九章算术〉及其刘徽注研究》序*

美国克莱因（M. Kline）的《古今数学思想》一书，在国际上被认为是最好的一本数学史专著。在该书作者自序的篇首，作者引用了 H. Poincaré 的一句名言：

> 如果我们想要预见数学的将来，适当的途径是研究这门科学的历史和现状。

笔者基本上同意该书作者对数学史的态度与观点，尤其同意所引用的 Poincaré 那句名言。遗憾的是，克莱因把数学的历史和现状实质上完全局限于西方的数学。为此，笔者认为为了澄清事实，对 Poincaré 的名言，应该补充一句作为注解：

特别是研究这门科学在中国的历史和现状。

出现于秦汉时代的《九章算术》与魏晋时期的《刘徽注》(以下简称《九章》与《刘注》)，是数学在中国最早、最完整的历史记录。《九章》与《刘注》，是研究数学在中国的历史和现状的钥匙。

克莱因一书的正确书名，事实上应该是《西方古今数学思想》。在全书 51 章中，只有标题为印度和阿拉伯的数学的第 9 章才涉及到非希腊传统的东方数学。尽管如此，在该章之首作者还提出下面的看法：

> 在数学史上，希腊人的后继者是印度人，虽然印度的

---

* 本文摘自《〈九章算术〉及其刘徽注研究》(李继闵著)，陕西人民教育出版社，1990 年。

数学只是在受到希腊数学成就的影响后才颇为可观。

只要对中国的传统数学略有所知，即知此语之谬。但是，我们不能轻以责己，而严以责人。造成这种局面的原因应该返求诸己。如果我们对自己数学的历史了解不多、认识不深，也不向西方的学者多作介绍，又如何能要求一位西方学者，克服文字上难以逾越的困难而对中国的传统数学在数学发展历史上的地位作出正确的评价。

作为一名中国的数学工作者，首先应对自己的数学历史有深刻的认识，为此必须首先对《九章》与《刘注》有确切的了解。

数系统的每一步完善都是数学进展的重要标志。无理数的发现，曾在西方引起了数学危机。负数与实数概念，则在西方很晚才得到确认。克莱因在《古今数学思想》中曾经说过：

> 负数虽然通过阿拉伯人的著作传到欧洲，但 16 世纪和 17 世纪的大多数数学家并不承认它们是数。

克莱因又说：

> 数学史上最使人惊奇的事实之一，是实数系的逻辑基础竟迟至 19 世纪后叶才建立起来。在那时以前，即使正负有理数与无理数的最简单性质也没有逻辑地建立，连这些数的定义也还没有。……这一事实说明数学的进展是怎样地不合逻辑。

然而在我国，远在《九章》之前就已有着举世无双的位值制十进制记数法。至迟在《九章》中，就已记载着有理数与正负数的各种运算规则。不仅如此，对于古代希腊认为迷惑不可理解的开根不尽之数，在《九章》与《刘注》中直截了当地"以面命之"，给出了独立成数的定义与某些运算法则。事实上，通过十进不尽小数的引入，以及开方与圆周率的极限计算，《九章》与《刘注》实际上已完成了整个实数系统。所谓

实数系统的严密逻辑基础，完全可以通过朴素的十进制小数来完成，而无需借助于19世纪才引入的Dedekind分割之类纡回曲折的概念。

数学研究现实世界中的数量关系与空间形式。在中国的传统数学中，数量关系与空间形式往往是形影不离并肩地发展着的。但在以欧几里得为代表的希腊传统里，则几何学独立于数量关系而以单纯研究空间形式的格局发展着。在《古今数学思想》中克莱因说：

> 代数虽在埃及人和巴比伦人开创时是立足于算术的，但希腊人却颠覆了这个基础而要求立足于几何。

希腊传统的这种排斥数量关系于几何之外的研究方式可能给数学包括几何带来了严重后果。在欧洲长时期黑暗的中世纪中，数学的发展陷于停顿，几何也是如此。笔者怀疑欧几里得那种单纯依靠艰涩而纡曲地进行的推理方式，正是造成这种停顿的重要原因之一。不论笔者的怀疑有多少真实性，一个无可否认的事实是：中世纪时阿拉伯世界，无疑是由于东方的影响，已经充分掌握了当时数量关系方面的许多知识与方法，可能还有不少自己的创造。通过回教、蒙古与土耳其的西侵，以及十字军的东征，这种知识与方法传入了欧洲，前面所说负数的传入正是其中之一。这种传入无疑促成了中世纪以后欧洲以数量关系为主而与欧几里得传统大相径庭的种种发明创造：小数，对数，符号，以至三次、四次方程的解法，等等。

与以欧几里得为代表的希腊传统相异，我国的传统数学在研究空间形式时着重于可以通过数量来表达的那种属性，几何问题也往往归结为代数问题来处理解决。面积、体积与圆周率的计算导致无理数概念的引入，相当于Cavalieri原理的刘祖原理的发现，以及极限方法的创立。把几何问题化为代数问题的做法，则导致方程、天元等概念的引入，多项式运算与消元方

法的建立，以及各种方程的系统解法，并使几何代数化有途可循，有法可依。17 世纪 Descartes 解析几何的发明，正是中国这种传统思想与方法在几百年停顿后的重现与继续。

《九章》与《刘注》，正是集中国这种传统思想与方法于大成、继往开来的传世杰作。

笔者曾在多种场合，指出我国的传统数学有它自己的体系与形式，有着它自身的发展途径与独到的思想体系，不能以西方数学的模式生搬硬套。我国的古代数学基本上遵循了一条从生产实践中提炼出数学问题，经过分析综合，形成概念与方法，并上升到理论阶段，精炼成极少数一般性原理，进一步应用于多种多样的不同问题。从问题而不是从公理出发，以解决问题而不是以推理论证为主旨，这与西方之以欧几里得几何为代表的所谓演绎体系旨趣迥异，途径亦殊。由于形形色色的问题往往归结为方程求解，因而方程求解就成为中国传统数学《九章》以来发展中的一条主线。这与西方数学之以定理求证为中心者正相对照。《九章》与《刘注》不仅提出了线性联立方程组的一般解法以及相伴而生的正负数概念，也给出了如何由开方与勾股类问题导致二次方程的范例。为此后千多年方程不断发展开其先河。诸如方程之类如何由简到繁，由特殊到一般，由具体到抽象的演变过程，从研读《九章》与《刘注》，可以得其梗概。

研究《九章》与《刘注》，不仅对于数学的历史，即使对于数学的现状，也可提高认识。

我国传统数学在从问题出发以解决问题为主旨的发展过程中建立了以构造性与机械化为其特色的算法体系，这与西方数学以欧几里得《几何原本》为代表的所谓公理化演绎体系正好遥遥相对。《九章》与《刘注》是这一机械化体系的代表作，与公理化体系的代表作欧几里得《几何原本》可谓东西辉映，

在数学发展的历史长河中，数学机械化算法体系与数学公理化演绎体系曾多次反复互为消长，交替成为数学发展中的主流。肇始于我国的这种机械化体系，在经过明代以来近几百年的相对消沉后，由于计算机的出现，已越来越为数学家所认识与重视，势将重新登上历史舞台。《九章》与《刘注》所贯穿的机械化思想，不仅曾经深刻影响了数学的历史进程，而且对数学的现状也正在发扬它日益显著的影响。它在进入 21 世纪后在数学中的地位，几乎可以预卜。

总之，要预见数学的将来，不能不研究《九章》与《刘注》所蕴含的深邃的思想在数学发展过程中的历史功绩，也不能不正视正在崭露头角的这种思想对数学现状的影响。本书对《九章》与《刘注》从各个角度作了全面的分析与介绍。书中不乏创见，对《九章》与《刘注》中无理数理论的阐发即是其中之一。笔者希望读者通过此书，不仅对中国传统数学在古今数学思想中的地位会有一个清晰的认识，而且能对数学将来的发展得出自己的结论。

在中外数学史讲习班开幕典礼上的讲话*

同志们：

今天是中外数学史讲习班的开幕典礼，我首先在此表示祝贺！我想，这种讲习班决不会只是办一次，以后可能每一年或至少每两年总要举行的。这不仅对于数学史的研究，而且对于整个数学的发展都是一件很重要的事！

了解历史，特别是数学史，正象江老已经说过的，它的重要意义是不言而喻的。假如你对数学的历史发展，对于一个领域的发生和发展，对一个理论的兴旺和衰落，对一个概念的来龙去脉，对一种重要思想的产生和影响等这许多历史因素都弄清了，我想，对数学就会了解得多，对数学的现状就会知道得更清楚、深刻，还可以对数学的未来起一种指导作用，也就是说，可以知道数学究竟应该按怎样的方向发展可以收到最大的效益。

这个讲习班，现在是取名为“中外数学史讲习班”。前面有“中”“外”两字，而且讲习的课程是分两部分，前面一部分是讲西洋的数学史，后面一部分是讲中国的数学史。我想，对数学史的现状来说，这样的做法也是比较合适的。当然，这是第一步，以后还应该有第二步、第三步。西洋数学史和中国

---

* 本文摘自《中国数学史论文集》（二）（吴文俊主编），山东教育出版社，1986年。

数学史分成两段来讲，是符合当前情况的，也是符合某些历史上的情况的。依我个人的理解，也不一定很正确，现在大学学的数学是西洋数学，是从希腊欧几里得系统下来的数学。从历史来看，我总觉得有两条发展路线，一条是从希腊欧几里得系统下来的，另一条是发源于中国，影响到印度，然后影响到世界的数学。这一条线现在不太显著，所以一讲到数学就是欧几里得统治下的、以演绎为主的公理化数学。可是，如果看看中国古代的数学就会知道它的味道、它的体系或者说它的思想和西洋数学是不一样的。我们中国的古代数学，看不出有什么公理、定理这一套的词，演绎体系也不是很明显的。可是，它自己有一套体系，这个体系也不是很清楚。但是我想可以借用严敦杰老先生的话来说明。严敦杰老先生在数学史上的成就是非常高的。在古代，中国的天文学和数学是不分家的，许多历法书里就有很多数学。严老先生对中国天文历法的研究也同样有非常高的成就。我理解严老有这样的想法：中国的天文发展与西方不一样，所以你必须从中国自己的角度来考虑中国自己的天文，然后才能够弄清中国的天文是怎么回事。你不能把西方的那个天文学生搬硬套到中国天文学的发展上来。* 我觉得严敦杰老先生对天文的这种想法也适用于中国的数学。当然，这要做大量的研究工作才能弄得清楚，而这在目前还不容易做到。总而言之，我觉得数学的发展有这么两个大的体系，大的思想，大的方向。它们是很不相同的。不过从利玛窦东来把西方数学输入之后，我们中国的那套思想就给排斥掉了。很少人能够懂得这一点，只有极少数数学史家知道不一样。我们的老前辈李俨先生、钱宝琮先生，还有现在健在的严敦杰先生，在

---

* 参看例如下文的最末一段，严敦杰，《中国古代的黄赤道差计算法》，载《科学史集刊》，1958年，第一期。

他们的著作里经常可以看到他们有这种意见和这种看法。可是很难把它系统整理成一套…（因转磁带，缺了一段）体系。

如果“公理化”是导源于希腊欧几里得的西方数学的主要思想，则我认为，我们中国的数学注重的“着眼点”就完全不一样。我用一个名称，叫“机械化”。“机械化”这个词过去是没有的，但现在它有特别的意义。现在是所谓“第三次浪潮”，是计算机、信息时代。我国古代数学的精髓是一种机械化的思想，一种机械化的方法，正好符合于现时代的要求和状况。因此，我觉得对中国古代的数学要特别加以重视。这一点在希腊欧几里得体系里是找不到的。我的这个看法是比较个别的，象江老说的是“一家之言”吧，这里提出来，希望今后研究数学史时能注意这一点。

刚才江老谈到一本数学史名著，即 M·克莱茵的《古今数学思想》。它着重在数学的思想，讲数学思想在历史上的发生、发展和影响。这本书是要把古今的数学思想介绍给读者，这些数学思想作者都是非常了解的。可是，作者讲它的书里不包括中国的数学，因为他没法儿懂。这还是比较实事求是的精神。书里还讲到许多印度数学，没有讲中国的数学。我想，这是数学史上这几十年应该弄清楚的一个问题。

在钱宝琮先生的《中国数学史》中，就提到中国与印度数学的关系。在英国，有一位数学史家叫 Kaye，他就明确地说过，印度数学是欠了中国数学的债的，印度的数学是从中国传去的。印度的数学有一个特点，它与西方的不一样，就是我说的机械化思想。这是占居主要的、统治地位的思想。现在，这种思想还不显著，可是也有些露头。最近这些日子，国外有一批学者来国内举办数学研究生教育研究中心，其中有一位对中国古代的数学史很重视。我以前不认识他本人，但我对他的老师很有兴趣。他的老师是印度人，写过一篇文章，登在美国数

学会月刊上，叫做《代数几何漫步》。代数几何是现在最难、最时髦而且认为是非常重要的一个领域。他的《代数几何漫步》中提出一种主张，这种主张用我们现在的话来讲，就是他不赞成古代希腊的那种做法。他是印度人，他就推崇印度人的那种作法，特别提到印度在十一二世纪的那个数学家 Bhaskara. 他很赞成的那种作法，实际上就是我刚才讲的中国古代的机械化。他认为，这种方法也能够取得西方欧几里得的那种成功。当然，他的主张大家都还不接受。他这么说，但是拿不出太多的东西来，他也做了一些工作，可是还不足以说明这种方法的优越性。但是，我觉得，他的主张是代表了某些人，他们已经注意到了这种机械化的方法有着潜在的巨大力量。我想，将来经过努力，是会发出芽来、开出花来的。当然，这是要经过很大的努力才行的。

这个讲习班第一次办，先介绍西洋数学史，然后再介绍中国数学史，这种方式是很合适的。第一步就是要把有关的知识普及给大家。然后，许多同志在学校里讲数学史的课，至少把现在已经有的成就普及到广大的青年一代，再下去，就要考虑中西方数学的交流与比较。我过去对数学史是有兴趣，偶尔翻翻，并没有认真下功夫。可是，我对某些书籍看了很气愤！比如说《数学简史》，已翻译成中文，作者斯特洛伊克是一个外国的共产党员。是一个有一定地位的数学家。他写这本书，应该说，有许多地方是写得很出色的。比如说，我非常赞成的一个地方，他说现在的中学教科书可以明显看出两个部分：一个部分是几何里面的欧几里得的希腊讲法，还有一部分算术、代数是东方色彩。如果把中学教科书拿出来看，这两个体系是贯穿在一起但并不是很调和的。也就是说，希腊的那个体系想在中学占统治地位，但它又是统治不了的。而且是根本不可能的。这是我的看法，也许武断了一些。当然，它还是很重要

的。斯特洛伊克在那本书中把这一点指出来是很出色的。可是，有的地方我是不满意的。比如，他对中国的数学也讲了一些。他提到两个人，一个是祖冲之，一个是秦九韶。他们非常重要而且成绩很大。可是祖冲之许多东西失传了，你就没办法说、没法全部理解它。斯特洛伊克提到祖冲之，提到 π 的计算，我觉得这一成就对数学的发展不是决定性的。中国数学具有重大意义的有的是，π 的计算不是最重要的。我认为，对圆周率来一个更精确的估计，这本身不是重要的。而得到这个结果的方法可能非常重要。但是它现在失传了，你也说不上是怎么一回事。秦九韶的"大衍求一术"也是这样的，它对数学的发展不是决定性的。有决定性意义的有的是，不能以他为代表。这还可以说斯特洛伊克是介绍了这两位数学家，他们的确重要，但我以为还不是中国数学的代表。然而，有一处引起我非常反感和气愤的是他讲《海岛算经》。在《海岛算经》中有一个公式，它是公元 3 世纪（263 年）得到的。那个时候，应该有东西方数学交流了。斯特洛伊克下了一个武断的结论，说这是从西方传过来的。我觉得这简直是荒谬透顶！这里有一种西方学者们牢不可破的"优越感"，即：我这里什么都有，你的肯定是从我西方传过去的。我想，他是受这样一种思想支配的。

前几年，有一位比利时的数学史家名叫莱伯热西特（U. Libbrecht），他到中国来座谈，我参加了。他讲过一些话，大意是：把希腊抬得那么高，是当时欧洲人要从落后状态中解放出来，需要有一种思想来鼓舞。他用了一个名称叫"欧洲中心论"（我是第一次听到），认为什么文化都是发源于欧洲的，你们东方嘛，都是"野蛮民族"，一切文化都是从我们欧洲传过去的。莱伯热西特提到这样一个名称，我觉得是相当中肯的。

我讲这一些，是要说明：我们讲数学史现在把中、西分开

来介绍，将来更重要的一步是要弄清楚东、西方数学的关系。东方数学和西方数学，正象斯特洛伊克那本书里讲的，是明显不同的两个体系，有不同的思想在里面。要说那么长的岁月里没有交流嘛，这是不可想象的。当然，不是象“欧洲中心论”和“西方至上论”的那些学者讲的，东方的东西是从西方传过去的。这是荒谬的。我们应该作为历史问题来考虑，应该实事求是，从我们掌握的资料来追查当时东方、西方学术上的交流是怎样的。一般说来，总是文化高的地区流向文化低的。我曾经拜访过北大历史系或是其他什么系的一位教授，是江老陪我去见的，他是回族人，谈到阿拉伯的数学，他就说过这么一句话：“水总是由高处向低处流的，文化一定是从文化高的地方传到文化低的地方。”应该是这样。这是从古代来讲的，现在东方落后了，当然应该接受西方先进的文化，引进他们的成就。所以，现在西方的学术流向我们中国是很自然的，只能说我们中国人自己不争气，落后了。不能不引进。而在古代是相反的，我们东方的文化是处在先进的地位，而欧洲是非常落后的。是一个宗教的“黑暗时代”，没有什么文化。十二、三世纪，他们甚至连加法都认为是学术上很难的东西，数学教科书上讲加法就很不错了。象这样落后的状况，你却说东方的文化不流向西方，而是西方的反而流到东方，这合理吗？当然，这是从“情理”方面来讲的，推测应该是这样的，查无实据。这个实据，我想应该是存在的，等待地下资料的发掘，这个发掘既需时日，也靠不住。我们不能把希望完全寄托在这上面。事实上，我相信在现有的资料里面，在我们大家所能看到的、能掌握的资料里，就可以分析出东方、西方交流的情况。这是要下功夫的事！这不光要看数学资料，还要看其他的资料。中国的数学很大一部分是包含在历代的历法即历书中，还有其他的需要借助的东西，下功夫不是一朝一夕的。而且你还要从第一

手材料里来，不能从第二手、第三手的材料中来。我们拿一本西方的数学史书来看，它们多得很，其中每一本数学史都是从希腊这样传下来的，那都是第二手、第三手的材料经过种种加工才写成的。你要真正了解历史，就必须追查第一手材料。但是，一开头就看第一手材料是不可能的、先要看第二手、第三手的材料去了解概况，再看第一手材料来检查。看第二手材料、第三手材料讲得对不对。从这方面来检查，这是要下功夫的。要看第一手材料，就要过文字关。少说一点，是希腊文、拉丁文罗。而你要真正弄清楚东、西方交流的历史，你就得掌握阿拉伯文、波斯文，懂得土耳其文，懂得这几个地区的文字。当然，现在我们是不可能做到的。可是，我想我们应该有志气来做！要说文字之难吧，中文大概是全世界文字之中最难的一种。可是你想，刚才我说的那位比利时数学家莱伯热西特，他写的一本《十三世纪的中国数学史》，就是分析、介绍秦九韶的《数书九章》的。他所引用的中文书之多，在我们国内虽不能说没有，至少也只是极个别的人才能看到它们。他的书后面还有一张表，把《数书九章》中的名词做了一个相当于英文的什么名词的工作。他读书是真正读懂了。又如，李约瑟的书中所引的中国书之多，我看中国人自己的书中也很少能看到那么多的。你看，外国学者就有这种志气，对于中文这么难的一种文字，他们就能学到手并且比大部分中国人学得好（我们有很多人还不太懂古汉语），这是不太简单的！

那么，中国人就没有一种志气、一种能力可以掌握阿拉伯文、土耳其文、中亚细亚各国的文字？我想，这是应该做到的。当然，这对某些人来说是做不到了。比如，江老就做不到了，我也做不到了。我想，白尚恕同志和主讲的几位也难做到了（当然，他们看的书是很多的），就是在坐的我看也是不太容易。但是，我想，经过这个讲习班，大家在各个高等学校开

设数学史这门课，培养出大批新生力量来，他们不仅是有条件而且有责任把东、西方的数学交流这个问题弄清楚。我们的讲习班不仅是在这里讲讲课，回去讲讲课，我认为还应该担负起某种责任来。这个责任，我们这一代事情那么多，不那么容易做到。但我想，许多同志回去开课的时候，应该使下一代把这个任务担当起来。要彻底把东、西方数学交流的问题弄清楚。这是能做到的！

今天，我拉拉扯扯说了一些，有一些是我个人的想法说得好听是“一家之见”，或者是一个人的“狂想”，随便说说。这次讲习班是第一次，以后继续干下去，中国数学史和西方数学史可以在全国各个高等院校普及，将来再进一步深入到两者历史上的关系，就可以呈现一种非常新的面貌！这个讲习班由北师大来主持，也不是很偶然的，因为北师大有历史传统。师大是搞数学教育的。数学教育与数学史是分不开的。数学史在北师大是向来受到重视的。现在有白尚恕同志，还有一些同志，特别是新任的北师大校长王梓坤同志他写过一本《科学发现纵横谈》里面有许多数学史与自然科学史的材料，深受欢迎，我想，数学史、科学史在北师大就有一个优越的环境了。我就讲到这里。

祝讲习班成功！

数学机械化

《可剖形在欧氏空间中的实现问题》的绪论*

一、实现或嵌入问题

依照 F.Klein 的经典的理论，几何学研究某种类型图象的某种类型的性质，而且也正由于所考虑的图象与性质各不相同，相应地形成了种种不同的几何学分支。如果细加分析，那么各种几何图象，尽管来源有别，归根到底往往归结为位于某一欧氏空间中的“具体”的图象，这个欧氏空间可以是有限维的，也可以是无限维的，即 Hilbert 空间。另一方面，为了要研究这些图象的内在的特性，也就是属于图象本身而与所在空间无关的那种特性，就有必要从一开始就以抽象而独立的形式来加以定义，例如从 Cantor 关于欧氏空间中点集的研究逐渐发展成的拓扑空间的概念，以及根据欧氏空间中光滑曲线、曲面等引伸而成的 Riemann 流形或微分流形的概念等。一个自然引起的问题是：如何能把“抽象”概念与“具体”事物恒同起来。这样一个问题的正面的答案我们称之为“实现”定理或“嵌入”定理，许多几何学中的基本定理正是属于这样一种性质。仅从拓扑学方面来说，就可以提到下面两个例子：

1．Urysohn 定理：Hilbert 空间中的子空间，即正则而有

---

* 本文摘自《可剖形在欧氏空间中的实现问题》（吴文俊著），科学出版社，1978 年。

可数基的那种拓扑空间。

2. Menger－Nöbeling 定理：欧氏空间中的闭囿集，即维数有限、有可数基而紧致的那种拓扑空间。

这些定理给与了某种类型的图象以一个总的概括。所谓空间的正则性、紧致性等等本来是对于欧氏空间中图象观察分析而抽象出来的，上面的两条定理说明，在某种类型图象所具有的无数的拓扑性质里面，这些性质是具有代表意义的，因此它们在拓扑中占据着重要地位乃是当然的。

对于这一类问题，我们还可进一步考虑，在一特定维数的欧氏空间中所能容纳的图象，例如平面中所有可能的线性图，三维空间中所有可能的 Riemann 曲面等问题。就拓扑学而论，我们将探讨以下问题。

问题Ⅰ（空间的拓扑实现）一个拓扑空间可实现为 N 维欧氏空间中的拓扑子空间的条件为何？

问题Ⅱ（复形的半线性实现）一个复形具有剖分可实现为一 N 维欧氏空间中的欧氏复形的条件为何？

问题Ⅲ（微分流形的微分实现）一个微分流形可实现为一 N 维欧氏空间中的光滑流形的条件为何？

这样的问题要求完全解决，自然希望是极少的。如果与 Riemann 流形在 N 维欧氏空间中的实现问题相比较，后者自 E.Cartan 以来即使在“局部实现”问题上成果也不多，由此就可以想到前述问题的困难程度如何。事实上，我们的工作也只限于提供一些方法，对上述问题Ⅰ—Ⅲ（主要是问题Ⅱ）求得部分成果而已。

这就是本书的目的。

二、已知的成果及其分析

在文献中曾经出现过研讨问题Ⅰ—Ⅲ的几种比较一般的方法，它们对于在 N 维欧氏空间中实现的可能性问题都曾提供了重要的知识，以下我们将逐一叙述这些方法中的三个方法。

方法Ⅰ 假设一个紧致空间 X 已实现为一 N 维欧氏空间 R^N 中的子空间，那么考虑 R^N-X 与 X 的性质间的关系，即将给出 X 可实现于这样一个 R^N 中的某些条件。这一方法可以追溯到 Alexander 的早期工作，例如，由 Alexander 对偶定理可知在整系数或 modz Betti 数间有以下关系：$p^i(R^N-X)=p^{N-i-1}(X)$，$i\neq 0$，$p^0(R^N-X)=p^{N-1}(X)+1$。由此即得 $P^N(X)=p^{-1}(R^N-X)=0$，而这给出了 X 可实现于 R^N 中的一个必要条件，特别是一个 n 维闭流形不能在 R^n 中实现。Hantzsche 应用了同样的方法，研究了当一个 n 维闭流形 X 实现于 R^{n+1} 中时，在 X 与 $R^{n+1}-X$ 的下同调群之间的对偶关系，由此推得了 X 可实现于 R^{n+1} 中时，用 Betti 数与挠系数来表达的一些必要条件。同样，在 X 是 n 维投影空间时，Hopf 通过对于 X 与 $R^{n+1}-X$ 间同调环的对偶关系以证明了一个 n 维投影空间不能在 R^{n+1} 中实现这一定理。与 Hopf 同样的想法，使 Thom 获得了更一般的结果：一个 n 维闭流形 X 实现于 R^{n+1} 中时用 X 的同调环表达的某些必要条件。近年来 Peterson 研究了 X 与 R^N-X（或更正确些说 S^N-X，这里 S^N 是假定 X 已在其中实现的一个 n 维球）间同调运算的关系，由此而获得若干有趣的实现定理。

方法Ⅱ 假设 X 已实现于 R^N 中，与第Ⅰ方法有所不同，我们可以不考虑 X 与其余集 R^N-X 的性质间的关系，而考虑 X 与它在 R^N 中的邻域的性质之间的关系，这个方法似乎是首

先由 Whitney 用之于微分流形在欧氏空间中的微分实现这一问题上的。依据这种思考 Whitney 又创立了球纤维丛理论，并引入了微分流形 M 的所谓 Stiefel－Whitney 示性类 $W^i\in H^i$（M，mod2），以及对偶 Whitney 示性类 $\overline{W}^i\in H^i$（M，mod2），这些类在拓扑学以及微分几何中都起了重要的作用。就仅与实现问题有关者而论，我们暂时可以提到下面已成经典的 Whitney 定理。

定理 1（Whitney） 一个 n 维微分流形 M^n，若能微分实现于一 N 维欧氏空间中，则必须有

$$\overline{W}^k\ (M^n)\ =0,\ k\geqslant N-n。\tag{1}$$

微分流形上 Stiefel－Whitney 示性类这一概念曾被 Pontrjagin 推广为更一般的示性类。其中除 Stiefel－Whitney 示性类外，最重要者为所谓 Pontrjagin 示性类 $P^{4k}\in H^{4k}$（M）与对偶 Pontrjagin 示性类 $\overline{P}^{4k}\in H^{4k}$（$M$）。与上面的 Whitney 定理同样，我们也有下面的

定理 2 一个 n 维微分流形 M^n，如果能微分实现于一 N 维欧氏空间 R^N 中，必须有

$$\overline{P}^{4k}\ (M^n)\ =0,\ 2k>N-n。\tag{2}$$

Thom 依据了同一原则但应用拓扑空间的拓扑实现，研究了实现于 R^N 中的空间 X 的 Steenrod 平方运算以及它在 R^N 中邻域的 Steenrod 平方运算间的关系，获得了下面的

定理 3 一个局部可缩紧 Hausdorff 空间 X，如果可在 R^N 中拓扑实现，必须有

$$Sm^iH^r\ (X,\ \mathrm{mod}\ 2)\ =0,\ 2i+r\geqslant N,\tag{3}$$

这里

$$Sm^i:\ H^r\ (X,\ \mathrm{mod}\ 2)\ \rightarrow H^{r+i}\ (X,\ \mathrm{mod}\ 2)$$

是由 Steenrod 平方运算从下关系所确定的某种同态：

$$\sum_{i+j=k} Sm^i Sq^j = \begin{cases} 0, & k>0\text{ 时}, \\ \text{恒同同态}, & k=0\text{ 时}。\end{cases} \qquad (4)$$

同一原则还曾被不少学者用以研究微分流形的微分实现问题，例如 Massey 研究了流形实现于欧氏空间中时它的“管形”的上同调环，Atiyah 则把这一管形作为所谓流形的 Grothenthick 环的一个元素来研究它，等等，似乎循着这一方向还有着广阔的前景。

方法Ⅲ　在 1932 年时，Van Kampen 发表了一篇非常有趣的论文，其内容是关于复形在欧氏空间中的半线性实现，而所用方法的原则则与前二者完全不同。由于它的重要性我们将在以下作较详细的描述。

很早以前就已经知道，一个 n 维单纯复形 K^n 必可在一 $2n+1$ 维欧氏空间中线性实现，Van Kampen 进一步探讨了 K^n 在 $2n$ 维欧氏空间 R^{2n} 中线性实现的问题并依下面的想法来进行：

设想把一个 n 维有限单纯复形 K^n “尽量可能”地线性实现于 R^{2n} 中。一般说来，我们自然不可能期望得到一个真正的实现，但将得到一个几乎近似于真正的实现。换言之，只有有限多个“奇点”的近似真正实现，这些奇点乃是 K^n 中两个本不相遇的 n 维单形放入 R^{2n} 中后可能有的交点。现在的问题是：(i) 从这些奇点的探讨是否可以导出一些只与 K^n 有关但与方法无关的不变量？(ii) 有什么方法以及在什么条件之下可以除去这些奇点以得到一个真正的（半线性）实现？

为此设 K^n 中的 k 维单形为 S_i^k. K^n 中两个没有公共顶点的单形将称为是分离的。命 A 为所有无序指数偶 (i, j) 的集合。对应于分离的 n 维单形 S_i^n 与 S_j^n 所成的偶，作一整数环上的向量空间 L，其维数恰为 A 中元素的个数。于是 L 中任一向量可表作一组整数 (a_{ij})，这里 $(i, j)\in A$. 对 K^n 中

每一对分离的单形 S_a^{n-1} 与 S_l^n 可作一 L 中的向量 $V_{la}=(a_{ij})$ 如下：如果 i，j 都$\neq l$，或其中之一，例如 $j=l$，但 S_a^{n-1} 非 S_i^n 的面时，即置 $a_{ij}=0$，否则置 $a_{il}=\pm 1$（符号适当地选取）。如果 $p-p'$是上面这种向量 V_{la}间的一个整系数线性组合，则 L 中的两个向量 p、p'将称为等价的。于是 L 中的向量分成了若干等价类，任一 K^n 到 R^{2n}中依前面所说的那种近似真正实现，将确定一 L 中的向量 $p=(a_{ij})$，其中 a_{ij}为 S_i^n 与 S_j^n 在确定定向 R^{2n}中的交截数。Van Kampen 的工作指出，不论 K^n 到 R^{2n}中的近似真正实现如何，相应向量 p 的等价类总是相同的，因之是复形 K^n 的一个不变量，而这给出了上面问题（I）的一个答案，记此不变量为 $V(K^n)$，则显然有下面的定理。

定理 4（Van Kampen） 一个 n 维有限单纯复形 K^n 如能在 R^{2n}中线性实现（或甚至半线性实现），必须有

$$V(K^n)=0 \tag{5}$$

（即等价类 $V(K^n)$ 含有 L 中的 0 向量）。

由 $V(K^n)$ 的定义可见由一 K^n 到 R^{2n}中的近似真正实现 f 所确定的向量 p，可以视为是 f 成为一个真正实现的某种“阻碍”，它体现了 f 与一真正实现之间差距的某种度量，因此自然地引起这样的猜测，即 $V(K^n)$ 之为 0 将使奇点的除去成为可能，以致条件（5）不仅是 K^n 可线性或半线性地实现于 R^{2n} 中的必要条件，同时也是充分条件。在 $n>2$ 时，Van Kampen 也曾发表了这样的一个证明，但其中一个根本性的错误使这一证明本身不能成立。这一问题因之虚悬直至晚近始获解决，这在下面还将论及。

如果 Van Kampen 只考虑了复形在欧氏空间中的线性或半线性实现的问题，那么 Whitney 与 Pontrjagin 在考虑微分流形在欧氏空间中的微分实现问题时，也使用了与之有某种类似的

"阻碍"方法。例如，早在1936年时，Whitney就证明了一个n维微分流形M^n必可微分实现于一$2n+1$维欧氏空间中。如果我们尝试把M^n尽可能地微分实现于某一N维欧氏空间中，而$N\leqslant 2n$时，一般说来将出现一些奇点，而象Pontrjagin所证明的那样，这些奇点将负载某些下闭链，它们的对偶上同调类将与合理地选择的近似真正实现无关而为流形不变量，在后来被称为流形的示性类，它们包括了在方法Ⅱ中已经提到过的Stiefel－Whitney示性类与Pontrjagin示性类。从这一点也可得出定理1与2来。再者，在$N=2n$时，M^n到R^{2n}中一个合理的近似真正实现将只有一些孤立的奇点，而除去这些奇点的一些方法使Whitney获得下面的且同样已成为经典的定理。

定理5（Whitney） 任一n维微分流形必可微分实现于R^{2n}中。

象由Shapiro与作者所指出的那样，在$n>2$时，这一Whitney的方法也可用在某些情况下移去一个n维复形到R^{2n}中一近似真正半线性实现中所出现的奇点，而这恰好弥补了Van Kampen原证中的缺陷，从而回答了在上述情形下的问题（Ⅱ）：

定理6（Van Kampen－Shapiro－吴文俊） 一个维数$n>2$的有限单纯复形K^n，如果有

$$V(K^n)=0,$$

则必可半线性地实现于R^{2n}中。

应该提到，象上面所描述的由Van Kampen与Whitney所建立的基本想法，已被Haefliger用之于微分流形微分实现问题上而且推广甚远，Haefliger有关流形的这一理论，在什么程度上可以推广到可剖形或复形的实现问题上，乃是一个极有意义的课题。

三、本书中的方法

本书中用以处理实现问题的方法，奠基于下面简单的观察：一个空间 X 到另一空间 Y 中的（拓扑）实现，需要存在一个一对一的 X 到 Y 中的映象 f，以致 $f(X)$ 在 f 下与 X 拓扑等价，这样一个实现映象 f，因之是一个"拓扑"的连续映象，其主要的特征之一是一对一的。因此所谓实现问题是一"拓扑性"的问题，而与流行于代数拓扑中大部分问题，主要是属于"同伦性"的问题有别。对后者，由于主要是"同伦性"的，故代数拓扑中的强有力的"同伦"方法，往往直接可以应用上去，而对前者则否。为了要克服这一潜在于实现问题中的困难，必须找到合适的方法把映象 f 的一对一性显露出来，然后再使用通常的同伦方法。

为此我们引入了空间 $\tilde{X}^*$ 与 X^*，各由 X 的所有有序与无序点偶 (x_1, x_2) 与 $[x_1, x_2]=[x_2, x_1]$ 所组成，这里 x_1、$x_2 \in X$，而 $x_1 \neq x_2$。试考虑从 Y 依同样方式导出的空间 $\tilde{Y}^*$ 与 Y^*，，于是一个实现映象 $F: X \to Y$ 将自然地引出实现映象 $\tilde{F}: \tilde{X}^* \to \tilde{Y}^*$ 与 $F: X^* \to Y^*$，这里 $\tilde{F}(x_1, x_2)=(f(x_1), f(x_2))$ 与 $F[x_1, x_2]=[f(x_1), f(x_2)]$。如果估计到映象 f 的一对一性在定义 $\tilde{F}$ 与 F 时即已用到，就不难引导到这样一个猜测：即使我们只考虑 $\tilde{F}$ 与 F 的连续性，也要比只考虑 f 的连续性者为多，而这仅仅是"同伦"方法在 $\tilde{F}$ 与 F 上的应用，这就可给出 X 到 Y 中实现可能性的一些实质性的线索。不出所料，真正的情况确是这样。

事实上，试记 $\tilde{X}^*$ 到 X^* 以及 $\tilde{Y}^*$ 到 Y^* 的自然投影各为 πx 与 πy，则有以下作为连续映象的可交换图象，其中 $\tilde{F}$ 与 F 的一对一性我们已置诸不顾：

$$\begin{array}{ccc} \widetilde{X}^* & \xrightarrow{\widetilde{F}} & \widetilde{Y}^* \\ \downarrow \pi X & & \downarrow \pi Y \\ X^* & \xrightarrow{F} & Y^* \end{array} \qquad (*)$$

现在的问题是：从这样一个可交换图象能得出些什么结论呢?

对于这样的问题是有现成的工具可用的。显然 $\widetilde{X}^*$ 是 X^* 上的一个二叶覆迭空间，其覆迭变换是一在 $\widetilde{X}^*$ 中没有定点而周期等于 2 的拓扑变换，于是 P. A. Smith 的著名理论指出对这样一对空间（$\widetilde{X}^*$，X^*）可配以一组上类 A^m（$\widetilde{X}^*$，X^*）$\in H^m$（X^*，$I_{(m)}$），$m \geqslant 0$，这里 $I_{(m)}$ 等于整数群或模 2 整数加法群，视 m 为偶数或奇数而定。由于 $\widetilde{X}^*$、X^* 与 πx 都完全依赖于空间 X，故这些类 A^m（$\widetilde{X}^*$，X^*）实际上是 X 自身的拓扑不变量而可合理地使用下面的记号

$$\Phi^m(X) = A^m(\widetilde{X}^*, X^*) \in H^m(X^*, I_{(m)})。$$

容易证明 Φ^m（X）$=0$，将蕴含不论 $i \geqslant 0$ 如何，都有 Φ^{m+i}（X）$=0$，因之可引进一整数 I（X）或 $+\infty$，使其为使 Φ^n（X）$=0$ 的最小整数。同样对 Y 也可定义 Φ^m（Y）与 I（Y），于是依 P. A. Smith 的理论从可交换图象（$*$）可得

$$F^* \Phi^m(Y) = \Phi^m(X)。$$

由此即得

定理 1 一个空间 X 如果能拓扑实现于一空间 Y 中，则必须有

$$I(X) \leqslant I(Y)。$$

特别，在 Y 等于一 N 维欧氏空间 R^N 时，可证 $I(R^N) = N - 1$。因之有以下定理作为推论。

定理 2 一个空间 X 如果可在 N 维欧氏空间 R^N 中拓扑实现，则必须有

$$I(X) \leqslant N-1,$$

或即

$$\Phi^N(X)=0。$$

这两定理虽然是 P. A. Smith 理论的直接推论，却是我们整个理论的核心。值得注意的是，这些定理，尽管如此不足道，却还包含了象 § 2 所叙述的 Whitney、Thom、Van Kampen 等人的主要结果作为它们的特款。更精确言之，我们有以下诸定理（ρ_2 表模 2 约化)。

定理 3　如果 X 是一局部可缩的紧 Hausdorff 空间，且有 $\rho_2\Phi^N(X)=0$，则有

$$Sm^iH^r(X, \bmod 2)=0, 2i+r \geqslant N。$$

定理 4　如果 M^n 是一 n 维微分流形，且有 $\rho_2\Phi^N$（M^n）$=0$，则

$$\overline{W}^k(M^n)=0, k \geqslant N-n \text{ 时。}$$

定理 5　一个 n 维有限单纯复形 K^n 的 Van Kampen 不变量 V（K^n)，如果适当地给以诠释，即为上类 Φ^{2n}（$|K^n|$)。

最后一个定理说明 n 维复形 K^n 的上类 Φ^{2n}（$|K^n|$）可以诠释为 K^n 到 R^{2n} 中线性或半线性实现的某种阻碍，对于其他的那些类 Φ^m（$|K^n|$）也同样，而这给出了这些 Φ－类的一个阻碍理论，这一理论由作者也由 Shapiro 所建立。特别是，象在 § 2 中及其末所叙述的那样，Van Kampen 的方法配合着 Whitney 的一个方法，足以移去 Van Kampen 原证中的一个缺陷，而与 § 2 的定理 5 相结合时，可用以证明下面的

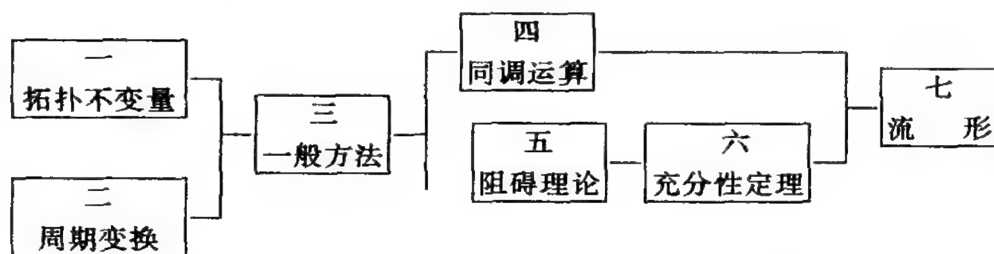
定理 6　一个维数 $n>2$ 的有限可剖形 K^n 在 R^{2n} 中可拓扑实现的充要条件是 Φ^{2n}（$|K^n|$）$=0$。

应该提到上面的简要理论可以循着种种不同的方向加以推广。首先，不仅可以考虑实现或嵌入问题，还可用同样方法来考虑局部实现或浸入以及同痕等问题，并可获得与定理 6 相当

的类似结果。其次，与空间 X 相配的空间 $\widetilde{X}^*$ 可视为由一两个点组成的标准空间 Z 到 X 中所有拓扑映象所构成的一个空间 $T(Z, X)$。于是不用这一 Z 以及 Z 到 X 中的拓扑映象作为标准空间与标准映象，我们也可使用其他的标准空间作为 Z，也可用其他类型比拓扑映象更灵活的映象来定义 $T(Z, X)$。例如在本书中我们即考虑了对任意质数 p，Z 是由 p 个点所成的空间以及 Z 等于碟形等这些情形。最后，在某些 $\Phi^m(X)=0$ 时并可发展一个二阶阻碍的理论，这在 X 等于微分流形的情形已证实是富有成果的，但由于这一理论正处于发展过程中，故本书不得不割爱。对此有兴趣的读者，可参看在《中国科学》1964 年度中发表的岳景中与作者的有关短文。

四、本书的结构

本书共分七章，其逻辑上的联系如下图所示：



第三章在全书中占据着中心的位置。它叙述了所使用的一般方法，说明了有关的基本概念，并提出了研究各类问题的基本不变量，即研究嵌入问题的 $\widetilde{\Phi}_p$-类，研究浸入问题的 $\widetilde{\Psi}_p$-类以及研究同痕的 $\widetilde{Ⓗ}$-类等。第一、二两章比较说来是属于准备性质的，前者指出了在第三章中所引进的那种基本不变量的特性，即拓扑不变但非同伦不变的那种性质，并叙述了构成具有这种特性的不变量的一个一般方法，后者则介绍了关于周期变换在没有奇点情形下的 P. A. Smith 的理论，它给出了工具

以构成对我们来说是重要的那些基本不变量。

最后四章中主要介绍第三章所述一般方法的应用。第四章引进了同调运算，与通常方法不同而仍然依据于 P. A. Smith 关于周期变换的理论，通过这些运算表达出嵌入与浸入的某些条件，特别是因此证明了§2 中 Thom 的定理 3 以及§3 中的定理 3。第五章包括了 Van Kampen 阻碍理论在嵌入、浸入与同痕方面的推广，依据这一阻碍理论将在第六章中证明，在一般情形时用我们的基本不变量来表达的那些必要条件，在某些临界情形下同时也是充分的。最后一章则致力于流形的嵌入、浸入与同痕问题，我们不仅证明了 Whitney 的那些定理（§2 的定理 1 与 2)，并证明了各种条件，即象第三章中用本书的基本不变量所表达的条件，象第四章中用同调运算表达的条件，以及在末一章中用示性类来表达的那些条件，在组合流形的情形下恰恰是彼此等价的。

应该提出，我们所考虑的问题主要是§Ⅰ的问题Ⅰ与Ⅱ，而问题Ⅲ实际上并未触及，即使是 Whitney 的定理以及各种条件的等价关系，我们也是就组合流形的拓扑嵌入与线性嵌入等等这些情形下来考虑的，而不是考虑微分流形的微分嵌入等等情形，因此整个处理方式可以说是初等性质的，唯一的例外或许是第六章的片段以及第七章的§5。事实上，微分流形的微分嵌入等问题需用的方法与工具和本书中所使用者相距甚远，也因为这一原因我们完全把它略去了。

在本书之末又附入了一些历史性注释，对于有关的历史发展以及与其他理论的联系作了一些说明。

作者非常感谢许多同志的帮助，特别应该提到北京大学的廖山涛教授与中国科学院数学研究所已故岳景中同志。对于前者，作者曾借用他不少重要的想法并贯穿于全书之中，而且在第七章中占着重要地位的吴振德的工作，是在廖先生的协助指

导下完成的；对于后者，作者曾与他有过不少可贵的讨论，而且第五章中一个重要定理由他给出了一个简单漂亮的证明，并得到他的允许而插入本书中。此外，作者还愿意提到李培信、江嘉禾、熊金城与虞言林等同志，他们都是在中国科学院数学研究所工作，与作者共事。

印刷电路与集成电路中的布线问题*

作为《可剖形在欧氏空间中的实现问题》示嵌类理论的一个实际应用，本文介绍了印刷电路与集成电路的布线问题。集成电路是 60 年代的产物，70 年代则发展为大规模的集成电路(LSI)，它的出现改变了电子计算机的设计思想。在应用分离元件的时代，一个主要的设计要求是力求采用最少量的元件与电路，而现在则已认识到主要矛盾在于 LSI 间的连接而非元件与电路的数量。丢开有关元件与电路图形的几何尺寸以及性能与工艺等因素不计，则连接问题或所谓布线问题在数学上可归结为一个线图在一个或几个平面中的嵌入与相连的问题。国外有些专著把这一问题视为未来新型计算机设计成败的关键，虽言之过甚，但自 60 年代初以至最近，国外每年都有不少文献探讨这一数学问题，这决不是偶然的。就作者查到的文献来说，国外远未解决这一已化成数学形式的问题，更不用说需要考虑性能工艺以及多快好省等其他许多因素了。

作者在 1966—1967 年间，从示嵌类理论入手，部分解决了上述问题，至 1973 年算是彻底解决了数学方面的问题，并脱离了示嵌类理论这一框框，所得结论与使用方法曾用简报形式发表于《科学通报》1974 年第 5 期 226—228 页，题目是《线性图的平面嵌入》。本文是这一简报的详细论述。欢迎有关

---

* 本文摘自《可剖形在欧氏空间中的实现问题》(吴文俊著)，科学出版社，1978 年。

同志把本文与国外类似文献进行分析比较，以资鉴别，并提出批评意见。由于作者接触实际少，更乏实践经验，同时写作时间仓促，在问题的提法与解法要求上与实际情况有所出入是很自然也是难免的，亟盼有关同志指出缺点错误，并提出新的问题，以便作更进一步的深入考虑。

一、问题的提出

1．问题的背景与来历

集成电路的出现产生了第三代计算机系统，而大规模集成电路的迅速发展又促进了第四代计算机的出现，印刷电路或集成电路以及大规模集成电路把大量元件或/与门电路构成于一块半导体片上。在这些元件或/与门电路间用导线相连，即所谓布线或互连。由于这些导线不能相互交叉，因此，即使在极简单的情形，要在一个平面上实现布线也往往是不可能的。为此，元件或/与门电路间的相互布线一般说来至少需要两层才行。这样就必须采用多层布线，利用通孔将不同两层上的布线连接起来，或则采取跨接与穿接的办法来避免不应有的交叉。

究竟元件或/与门电路应如何配置，各层上的导线应如何布设，不同层间应如何通孔连接，或应在何处跨接穿接，即现在所称布线问题的主要内容。自然，在这里我们将元件或/与门电路的形状大小，以及导线的粗细长短与是否过密等这一类内容已略去不加考虑，更不用说牵涉到电路性能这一类内容了。

作为一个简单的实例说明，试考虑图 1 所示一个通用的触发器电路。电路中有：

晶体管 2（$T1—2$）；

二极管 8（$D1—8$）；

电　阻 10（$R1$—10）；

电　容 4（$C1$—2）；

以及电阻二极管元件组 4（$DR1$—4）。

各端⓪指地，①—③为固定电源，⑤，⑥为输出端，Q 与 $\overline{Q}$，CP 为打入端，S 与 R 各为置位与复位端。

显然，这样一个电路要配置在同一平面层上是不可能的。由于电源与地线一般布设在与逻辑线不同的平面层上，因而可把这些端线以及直接相连的元件略去而只考虑所得如图 2 所示简化后的电路即可。

粗略看来，如果我们将原给电路图 1 或简化后图 2 中的所有横向导线布设在一平面上，所有纵向导线布设在另一平面上，再按导线相接处用通孔相连，则理论上任何复杂的电路图只需用两个（或三个）平面层就可完成布线了。但这样通孔数目将不必要地过多，而通孔的存在正是使集成电路性能与成品率大为降低的重要因素之一。因之这种做法不能符合实际需求而需另觅途径。这正是问题之困难所在。

仍回到前面图 2。注意各输入输出端需配置在平面层的边缘上，又在彼此相连且无元件阻隔的导线上电位到处相同，这些电位在图中记作 $V4$—14（在原电位图 1 中电位为 $V0$—18）。由于不考虑元件的几何形状与大小，我们可把这一简化电路图改成图 3 那样的线图，其中各元件与电位都用点表示，B 点则代表边缘。容易看出，简化电路的布线问题与这一线图的布线问题完全等价。解决了后一问题，即可获得原来问题的解决方法。较详细的说明可参阅例如“Goldstein-Schweikert，Bell System Technical Journal，52（1973），135”。

显然，即使象图 3 那样简单的线图，要布设在同一平面层中也是不可能的。为了避免出现不应有的交叉，必须放弃若干连线，把它们布设在另一个（必要时另几个）平面层上，再用

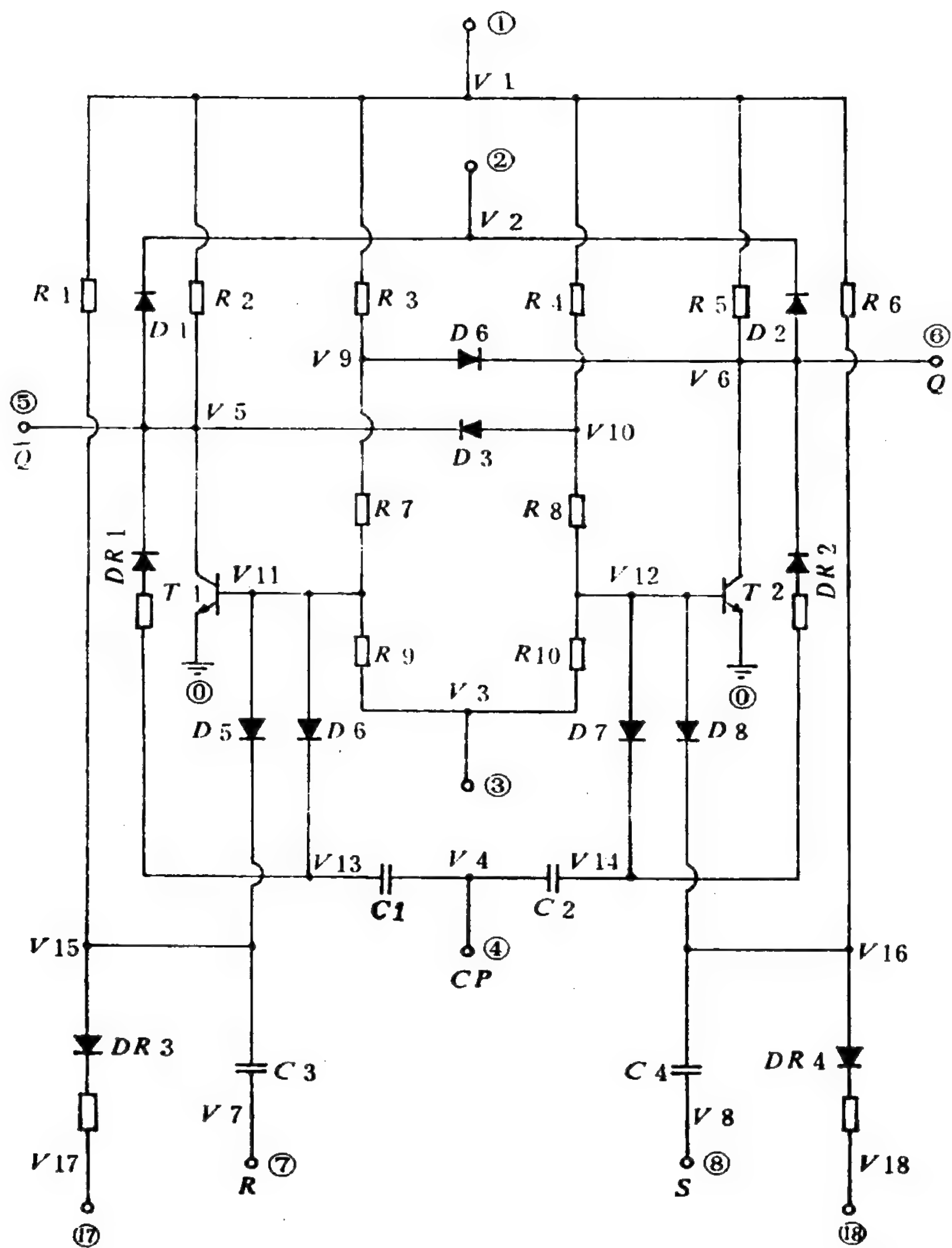

①
V 1
②
V 2
R 1
R 2
R 3
R 4
R 5
R 6
D 1
D 6
D 2
V 9
V 6
⑥
Q
⑤
Q̄
V 5
V 10
D 3
R 7
R 8
DR 1
T 1
V11
V12
T 2
DR 2
R 9
R10
⓪
V 3
⓪
D 5
D 6
D 7
D 8
③
V13
V 4
V14
C1
C 2
V15
④
CP
V16
DR 3
C 3
C 4
DR 4
V 7
V 8
V17
⑦
R
⑧
S
V18
⑰
⑱

图 1

通孔或跨接穿接与其它部分连上。但究竟应放弃哪些连线，使余下的部分可布设在同一平面层上，且其具体的布设方法如何，致使通孔或跨接穿接的数目尽量减少，则决不是一个简单的问题。象上面这种简单的实例，固然不难通过尝试与错误（trial and error）的办法来做出布线方案，但对于一个包括成百上千个元件或/与门电路甚至更多的复杂情形，则如果没有一个切实可行的方法指导系统地进行布线，盲目的尝试将是不可能的。就作者所知，至少从国外公开发表的资料看来，文献虽多，但问题远没有解决。

本附录的目的，在于给出彻底解决上述布线问题的一个切实可行的方法，所用通孔数目也将大量减少，虽不能保证达到最小限度。

如所周知，多层布线技术是大规模集成电路中较关键的技术，导线布设是印刷电路板设计最中心的问题，大规模集成电路技术的最后成败将取决于这方面的进展和革新。而且，由于大规模集成电路互连的复杂性，还需要一定的自动布线方法。自动化布线问题的解决，将使一台大型计算机的设计工作，只要把一些必需的逻辑方程式送入一台较小的计算机，就可坐而待成，而且能高速度地把设计方案传送出来。我们认为，本附录不仅提供了这样做的原则与方法，而且这个方法是切实可行的。我们希望在不久将来看到这个方法的具体实施。

2. 问题的数学形式

从上节已知集成电路的布线问题可化为把线图布线于平面中的数学问题，本节将这一数学问题以较明确的形式重述于下，首先说明用到的一些词汇与符号。

所谓一个线图 G，是指一些点（称为顶点）以及这些顶点之间的某些连线（称为棱，所连的顶点称为棱的端点）所构

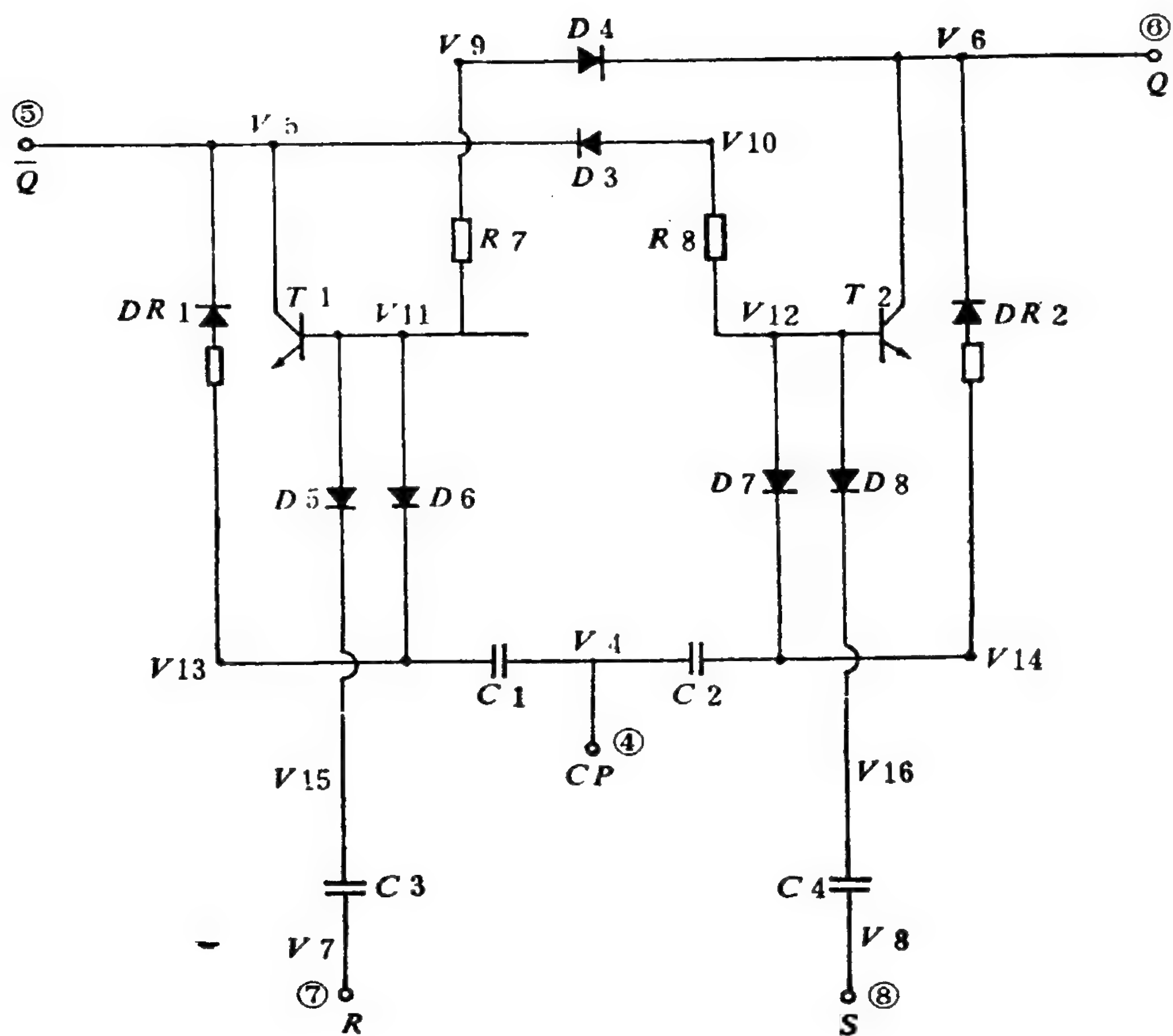

D4
V9
V6
⑥
Q
⑤
Q̄
V5
D3
V10
R7
R8
T1
V11
V12
T2
DR1
DR2
D5
D6
D7
D8
V13
V4
V14
C1
C2
④
CP
V15
V16
C3
C4
V7
V8
⑦
R
⑧
S

图 2

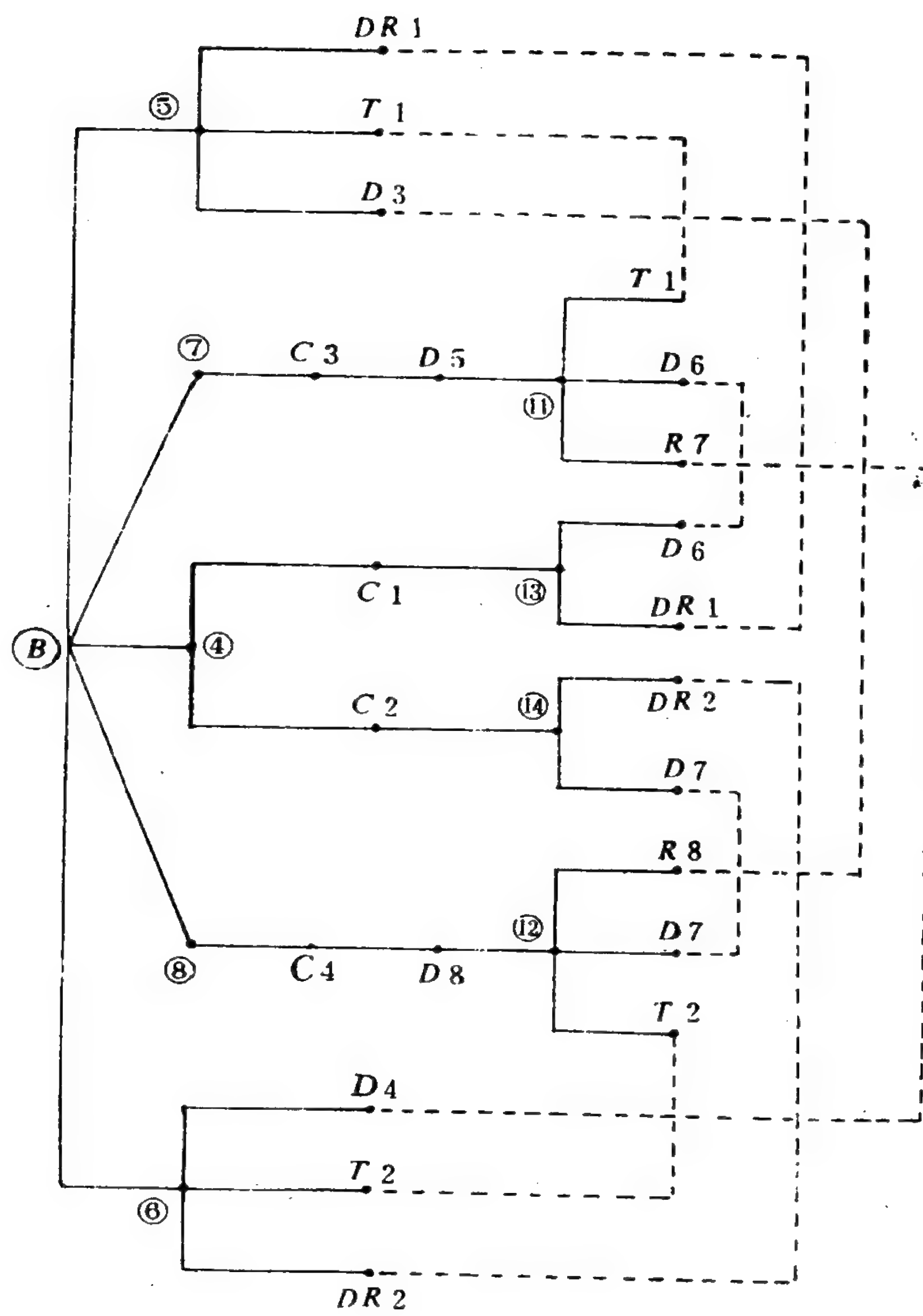

DR 1
⑤
T 1
D 3
T 1
⑦
C 3
D 5
D 6
⑪
R 7
D 6
C 1
⑬
DR 1
B
④
C 2
⑭
DR 2
D 7
R 8
⑫
D 7
⑧
C4
D 8
T 2
D4
T 2
⑥
DR 2

图 3

成的图形。在试图将 G 画入平面中时，规定每一棱都画成一个线段或一个不自交也不首尾相接的折线，例如，一个棱可画成下面的形状：

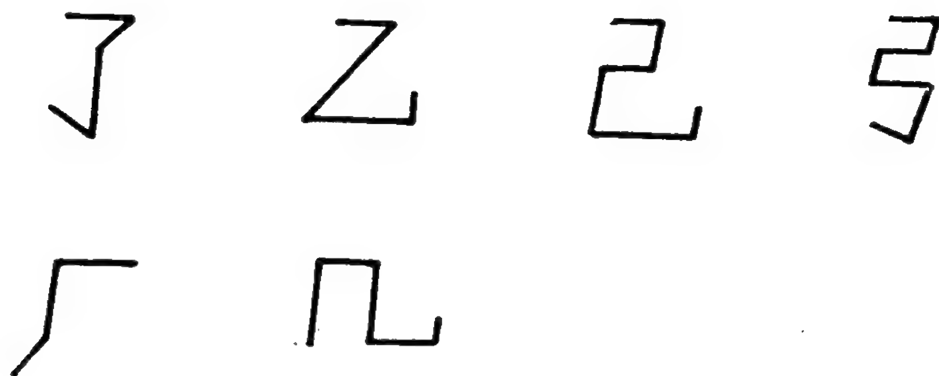


图 4

我们规定，将 G 画入平面时限于下面的方式：对 G 中任意两棱，画入平面后的折线，除原来可能有公共端点之外，只能在相遇处相截，即相遇之处不在折线的转折处，而在折线中段，且在该相遇处相交成角。记画入的方式为 f，则称 f 为 G 在平面中的一个浸入，记作 $f: G \subset\!\!\!\cdot$ 平面。

所谓线图 G 在平面中布线，是指将 G 如实地画在平面之中，即不相遇（或只有端点公共）的棱画入后的折线也不相遇（或只有相应端点公共）。这一般是不可能的。如果可能，我们就说 G 能在平面中布线，或 G 能嵌入平面，或说 G 具有平面性。仍把画入的方式记作 f，则称 f 是 G 在平面中的一个嵌入，并用记号 $f: G \subset$ 平面来表示。显然嵌入也是一个浸入，但浸入不必是嵌入。

如果 T 是 G 的一个子图，在把 G 画入平面成为一个浸入时，对 T 来说是一个嵌入，但对于在 G 中而不在 T 中的那些棱来说，画入后至多除原有可能的公共端点外都不再与 T 相遇，则称这一画法 f 为相对于 T 的浸入，记作 $f: G \underset{T}{\subset\!\!\!\cdot}$ 平面。



图 5

设线图 G 具有平面性而 f，g 是两个嵌入。如果在 f，g 下各棱画成的折线形状长短各不相同，但相互位置关系相仿，则我们将迳直视 f，g 为相同的嵌入而不加区别。例如，有四个顶点 0，1，2，3 以及三个棱 01，02，03 的入字形线图，可以嵌入成如图 5 所示的各种图形。这些嵌入都可视为与左下面的嵌入相同：

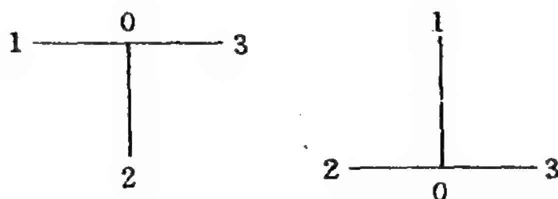


图 6

但如果把以上各图中的 1，2 两顶点互易，则所得诸嵌入彼此相同，而与以前诸嵌入相异，易知入字形线图的不同嵌入恰有图 6 的两种。

设 T 是 G 的子图而 f，g 是 G 相对于 T 的两个浸入，如果在上面意义下 f，g 限制在 T 的嵌入是相同的，则称 f，g 相对于 T 是相同的浸入。

若 f，g 是 G 的相同嵌入，则记作 $f \sim g$。若 f，g 相对于 T 是相同的浸入，则记作 $f \sim g$。

就本附录有关的布线或嵌入问题来说，不妨对线图 G 作以下限制，以后不再交待。

限制（1）G 是连通的，即 G 不能分解成两个没有顶点公共的子图，或即对任两顶点在 G 中至少有首尾相继的一组棱构成一个连接这两顶点的通道。

限制（2）G 不能分解成两个只有一个顶点公共的子图。

限制（3）G 中连接两个顶点的棱至多只有一个。

于是，以印刷电路或集成电路布线为背景的数学问题可概括为关于线图的以下几个问题：

（1）平面化问题。

从线图 G 的诸棱中挑出一部分（称为非平面性部分），使所余部分所成线图 G'（称为平面性部分）可布线，即嵌入于平面中。

（2）具体布线或具体嵌入问题。

对于已知可嵌入于同一平面中即已知有平面性的线图 G'，给出在平面中具体嵌入的方法。

（3）布线或嵌入分类问题。

对能嵌入的线图，将所有可能的不同嵌入方法列举出来进行分类。

为解决以上诸问题，须先在理论上给出一个线图可在平面上布线或具有平面性的判准。这样的判准还须是切实可行的，即可表达成带有程序性的算法，以致所有挑出非平面性部分以及将所余部分具体布线的方法都可编成程序，因而得以在电子计算机上给出具体方案。最后这一步将称为机器化问题。

上述整个问题可表达成下图所概括的几个部分，解决办法见以后各节。

诸问题的解决途径约略可说明如下。

一个没有回路（也称闭圈）的线图称为一个树形。树形之能嵌入平面是显然的，而一个一般线图 G 之所以不能嵌入平面主要是由于 G 中有着众多回路所引起，因之我们的方法在于在 G 中任意选择一个尽可能大的树形 T，称 T 中的棱为树棱，T 外的棱为外棱，先将 T 用例如方式 f 任意嵌入平面，再将诸外棱逐一添入以考查是否能扩充为整个 G 的嵌入。必要时放弃若干外棱并改变 T 的原来嵌入 f，以使扩充成为可能。

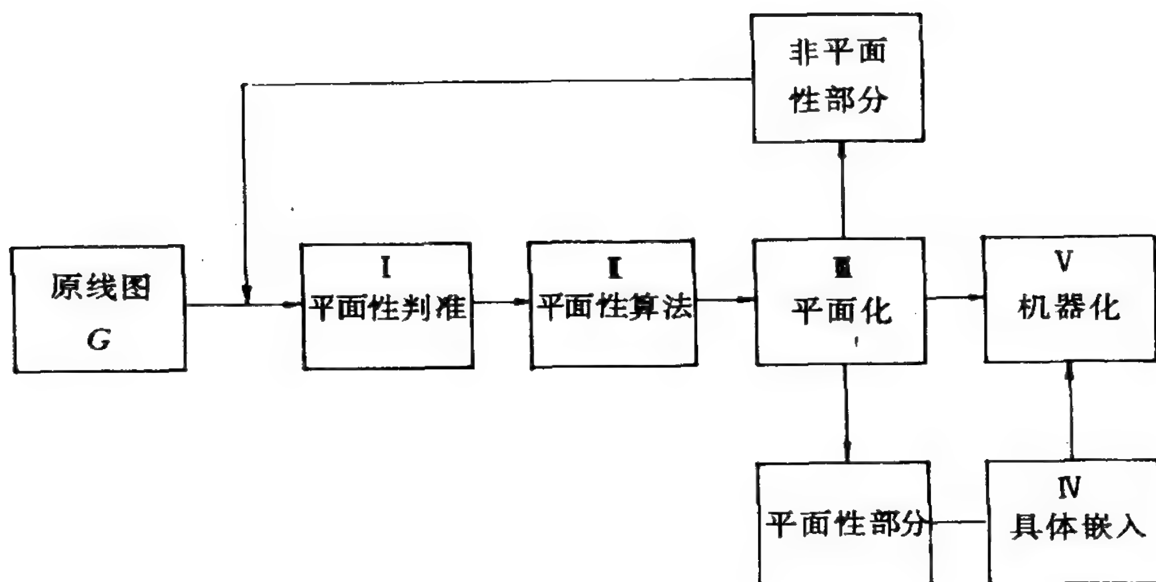


图 7

在第二部分中我们对树形的嵌入进行了分类。第三部分分析了一般线图 G 在所择树形 T 已嵌入后扩充为 G 的嵌入时所遇到的矛盾，由此获得 G 具有平面性的判准与算法，并给出了图 7 中问题 Ⅰ，Ⅱ，Ⅲ 的解答。第四部分则给出将所余平面性部分嵌入平面的具体作法，即解答问题Ⅳ。最后，第五部分列举了所有可能的具体嵌入。

我们的解决方法都切实可行且带有程序性，因而可据以机器化。其中一部分已由数学研究所吴方同志帮助编成程序并在 DJS－21 机上试过，其余部分则尚在编试中。

二、树形的嵌入问题

1. 树形的嵌入

问题 若 T 是一树形，则 T 必可嵌入平面，但可有多种不同的嵌入方法。例如一个 H 图形可如图 8 有 4 种不同的嵌

入法，依次记为 f^1，f^2，f^3，f^4：

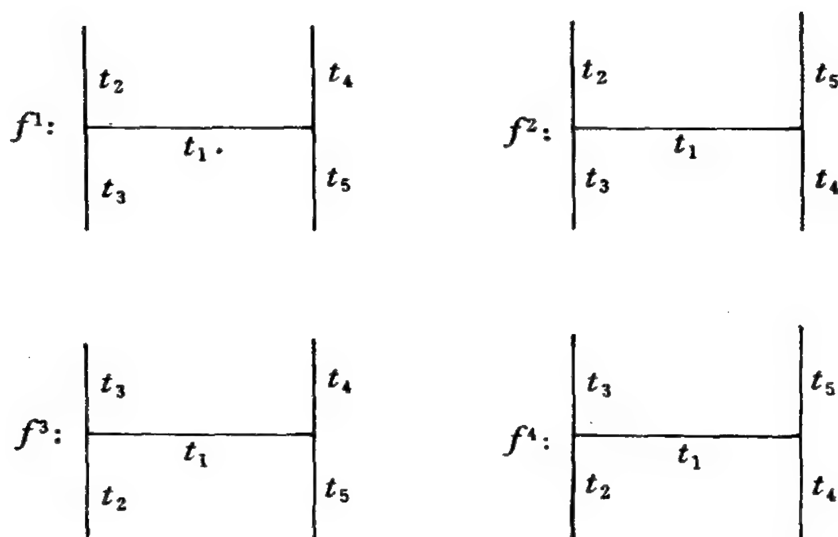


图 8

又如一个 X 图形如图 9 有 6 种不同的嵌入法，依次记为 g^1，…，g^6。

问对任一树形 T，如何确定它的不同嵌入方法。

剖析 考虑 T 的任一顶点 v 与以 v 为端点的任三（依一定次序排列的）树棱 t_i，t_j，t_k。这称之为一个三棱组，记作 $\langle t_i, t_j, t_k\rangle$。例如在 H 图形中，有三棱组 $\langle t_1, t_2, t_3\rangle$ 与 $\langle t_1, t_4, t_5\rangle$，在 X 图形中有三棱组 $\langle t_1, t_2, t_3\rangle$，$\langle t_1, t_2, t_4\rangle$，$\langle t_1, t_3, t_4\rangle$，$\langle t_2, t_3, t_4\rangle$。其余的三棱组都由这些组中三个棱改变排列次序而得，可以不再考虑。

给定一个嵌入 f: $T\subset$平面。一个三棱组 $\langle t_i, t_j, t_k\rangle$ 的三个棱 t_i，t_j，t_k 在 f 下绕 v 旋转的次序可以是反钟向也可以是顺钟向的，例如对 H 图形的嵌入 f^1，…，f^4，各三棱组旋向的顺反可列表如下：

H 图形三棱组旋向顺反表

三棱组 \ 嵌入	f^1	f^2	f^3	f^4
$\langle t_1, t_2, t_3\rangle$	反	反	顺	顺
$\langle t_1, t_4, t_5\rangle$	顺	反	顺	反

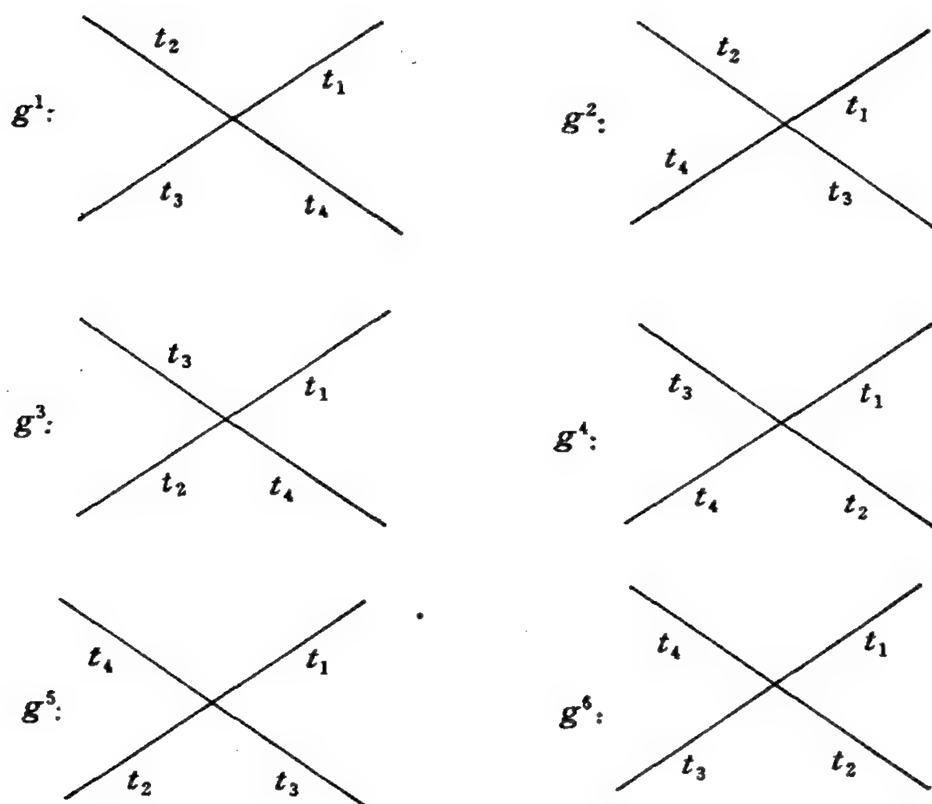


图 9

同样对 *X* 图形有下表

显然两个嵌入 $f: T\subset$平面与 $f': T\subset$平面是否相同可由各三棱组在 f 下与在 f' 下的旋向是否相同来加以区别。由于旋向恰有顺反两种不同情形，故可用一个模 2 数（0 或 1）来标

志其不同。为此，对任一嵌入 $f: T \subset$ 平面与三棱组 $\langle t_i, t_j, t_k \rangle$ 引入一旋向数 f_{ijk}，使

X 图形三棱组旋向顺反表

三棱组 \ 嵌入	g^1	g^2	g^3	g^4	g^5	g^6
$\langle t_1, t_2, t_3 \rangle$	反	反	顺	顺	反	顺
$\langle t_1, t_2, t_4 \rangle$	反	反	反	顺	顺	顺
$\langle t_1, t_3, t_4 \rangle$	反	顺	反	反	顺	顺
$\langle t_2, t_3, t_4 \rangle$	反	顺	顺	反	反	顺

$$f_{ijk} = \begin{cases} 0, & t_i,\ t_j,\ t_k \text{ 在 } f \text{ 下依反钟向旋转时,} \\ 1, & t_i,\ t_j,\ t_k \text{ 在 } f \text{ 下依顺钟向旋转时。} \end{cases}$$

对 (i, j, k) 的一个排列 $P = \begin{pmatrix} i & j & k \\ i' & j' & k' \end{pmatrix}$，显然有

$$f_{i'j'k'} = f_{ijk} + \varepsilon_p,$$

这里

$$\varepsilon_p = \begin{cases} 0, & p \text{ 是偶排列时,} \\ 1, & p \text{ 是奇排列时。} \end{cases}$$

对于所有的三棱组 $\langle t_i, t_j, t_k \rangle$，这些相应的旋向数或简称旋数 f_{ijk} 构成一个模 2 数组。在 H 图形为（相差一排列的诸三棱组只保留其中一个，下同）：

$$(f_{123},\ f_{145}).$$

对于嵌于 $f^1, \cdots, f^4$，这一数组依次为

$$(0,\ 1),\ (0,\ 0),\ (1,\ 1),\ (1,\ 0).$$

同样，对 X 图形在嵌入 $g: X \subset$ 平面下有一相应旋数组

$$(g_{123}, g_{124}, g_{134}, g_{234}).$$

对于嵌入 $g^1, \cdots, g^6$，这一数组依次为

$$(0, 0, 0, 0), (0, 0, 1, 1),$$
$$(1, 0, 0, 1), (1, 1, 0, 0),$$
$$(0, 1, 1, 0), (1, 1, 1, 1)。$$

显然，对于一个确定的树形 T，不同的嵌入将有不同的旋数组。例如 H 图形的不同嵌入恰与不同的旋数组相对应，其种数为 $2^2=4$。但并不是每一个数组都是某一嵌入的旋数组，例如对 X 图形，数组($g_{123}, g_{124}, g_{134}, g_{234}$)可取 $2^4=16$ 种不同的组值，但实际上只有 6 种组值才能成为嵌入的相应旋数组。这说明这些旋数间必须满足某种制约关系。从上表可知，一个关系是

$$g_{123}+g_{124}+g_{134}+g_{234}=0$$

或

$$g_{123}=g_{124}+g_{134}+g_{234}。$$

因而 g_{123} 由 g_{124}，g_{134}，g_{234} 完全确定。这反应了如下事实：把 t_4 的位置固定，则 t_1，t_2，t_3 中任两棱与 t_4 的旋向如果已定，即 g_{124}，g_{134}，g_{234} 的值已定时，t_1，t_2，t_3 的旋向完全确定，即 $g_{12}{}^3$ 的值也因之而定。但除了上述关系外，（g_{124}，g_{134}，g_{234}）的值也不是完全任意的，因为它们有 $2^3=8$ 种不同的组值，而嵌入的方法只有 6 种。这说明 g_{124}，g_{134}，g_{234} 还有其他必须满足的关系。在解决这一问题之前，我们先引入树根的概念，以使问题简化如下。

任取树形 T 的一个自由端 O，称之为树根，于是任一不同于 O 的顶点 v，在树形中自 v 至 O 恰有一通道，记之为 P_v。在通道 P_v 上以 v 为端点的树棱记为 t_v。设 t_i，t_j，t_k 为 T 中以 v 为前端（即自 O 出发到 t_i 时先到达的端点）的任意三棱（图 10），由前知，f_{ijk} 由 f_{vij}，f_{vik}，f_{vjk} 完全决定：

$$f_{ijk}=f_{vij}+f_{vik}+f_{vjk}。$$

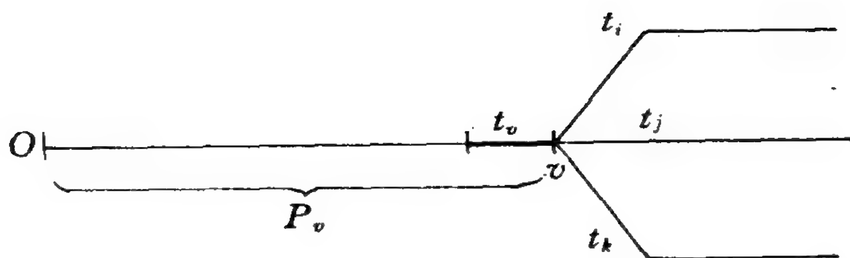


图 10

因而对于刻划树形的嵌入，可易数组（f_{ijk}）为个数较少的数组（f_{vij}）。为此，我们将 f_{vij} 简化记号为

$$f_{vij} = f_{ij},$$

这里

$$f_{ij} = f_{ji} + 1,$$

且

$$f_{ij} = \begin{cases} 0, t_v, t_i, t_j \text{ 在 } f \text{ 下反钟向排列时}, \\ 1, t_v, t_i, t_j \text{ 在 } f \text{ 下顺钟向排列时}。 \end{cases}$$

仍称 f_{ij} 为旋数(相对于选定的树根 O 而言)。这些旋数构成一个旋数组（f_{ij}），在组中每一对以同一顶点作为前端的两个树棱 t_i，t_j 保留了一个旋数 f_{ij} 或 f_{ji}。例如，在 H 图中取 t_2 的自由端为树根 O，在 X 图中取 t_1 的自由端为树根 O，则对嵌入 $f:H\subset$平面与 $g:X\subset$平面的旋数组可各取为

$$(f_{13}, f_{45})\text{与}(g_{23}, g_{24}, g_{34})。$$

综合上述,于是有下面的结论。

嵌树术 对于树根为 O 的树形 T,任一嵌入 f 确定一组旋数(f_{ij})。两个嵌入 $f,g:T\subset$平面是否相同可视相应旋数组(f_{ij})与(g_{ij})的完全相同与否而定:

$$\boxed{f \sim g \Leftrightarrow (f_{ij}) = (g_{ij})}$$

以下假定 T 的树根 O 已选定,所有旋数均对选定的树根

O 而言。

2．旋数关系（特殊情形）

问题 设 t_i，t_j，t_k为以同一顶点 v 为前端的任三树棱，由于 f_{ij}，f_{jk}，f_{ki}的每一数均可为 0 或 1，故数组（f_{ij}，f_{jk}，f_{ki}）可有 8 种不同的组值，但 t_i，t_j，t_k在f 下的不同旋向只有 6 种，因而（f_{ij}，f_{jk}，f_{ki}）的 8 种组值中至少有两种实际上不可能出现。换言之，在 f_{ij}，f_{jk}，f_{ki}间必然存在某种关系作为制约，试求这种关系的具体表达形式。

（特殊）旋数术 若树棱t_i, t_j, t_k以同一顶点为前端，则置

$$Q_{ijk}(f) = f_{ij}f_{ik} + f_{jk}f_{ji} + f_{ki}f_{kj}$$

时，必有

$$\boxed{Q_{ijk}(f) = 1}$$

且满足这一关系的不同数组(f_{ij}, f_{jk}, f_{ki})恰有 6 个，各与 t_i, t_j，t_k 在平面中 6 个不同的确定旋向相对应。

注释 如前记自 v 至树根 O 通道 P_v 上以 v 为端点的树棱为 t_v。在嵌入 f 下自 t_v 起作反钟向旋转时 t_i, t_j, t_k 的次序可有 6 种不同方式，记其相应 f_{ij}等值列作下表。

情　形	f 下反钟向次序	f_{ij}	f_{jk}	f_{ki}
(1)	t_v, t_i, t_j, t_k	0	0	1
(2)	t_v, t_i, t_k, t_j	0	1	1
(3)	t_v, t_j, t_k, t_i	1	0	0
(4)	t_v, t_j, t_i, t_k	1	0	1
(5)	t_v, t_k, t_i, t_j	0	1	0
(6)	t_v, t_k, t_j, t_j	1	1	0

从表中可看出，$(f_{ij}, {}_{jk}, f_{ki})$不能取的数组为

情　形	f 下反钟向次序	f_{ij}	f_{jk}	f_{ki}
(7)	不可能出现	0	0	0
(8)	不可能出现	1	1	1

今尝试作 f_{ij}, f_{jk}, f_{ki}的代数式 F(系数是模 2 数)，使对前 6 组值都有 $F=1$ 而对后二不可能取的数组有 $F=0$。显然 F 不能取为f_{ij}, f_{jk}, f_{ki}的一次式。其次对二次式进行尝试，此时二次项

$$f_{ij}^2=f_{ij}, f_{jk}^2=f_{jk}, f_{ki}^2=f_{ki},$$

而对上 8 个数组的二次项 $f_{ij}f_{ik}$等值依次可列成下表：

情　形	f_{ij}^2	f_{jk}^2	f_{ki}^2	$f_{ij}f_{ik}$	$f_{jk}f_{ji}$	$f_{ki}f_{kj}$
(1)	0	0	1	0	0	1
(2)	0	1	1	0	1	0
(3)	1	0	0	1	0	0
(4)	1	0	1	0	0	1
(5)	0	1	0	0	1	0
(6)	1	1	0	1	0	0
(7)	0	0	0	0	0	0
(8)	1	1	1	0	0	0

由表知合于条件的一个代数式是

$$f_{ij}f_{ik}+f_{jk}f_{ji}+f_{ki}f_{kj}。$$

术文中的 $Q_{ijk}(f)$即由此而来。据以上分析，自然得出术文。

3. 旋数关系(一般情形)

(一般)旋数术　设 $t_1, \cdots, t_s$ 为以同一顶点 v 为前端的任意树棱。由前(特殊)旋数术知,对其中任意三棱 t_i, t_j, t_k 在嵌入 $f:T\subset$平面下应有

$$Q_{ijk}(f) \equiv f_{ij}f_{ik} + f_{jk}f_{ji} + f_{ki}f_{kj} = 1。$$

换言之,引入一组(模 2)未知数($i, j = 1, \cdots, s; i \neq j$)

$$\varphi_{ij} = \varphi_{ji} + 1,$$

并置($i, j, k = 1, \cdots, s$ 且互不等)

$$Q_{ijk}(\varphi) = \varphi_{ij}\varphi_{ik} + \varphi_{jk}\varphi_{ji} + \varphi_{ki}\varphi_{kj},$$

则数组(f_{ij})应是方程组

$$(Q) \qquad Q_{ijk}(\varphi) = 1$$

的一个解答。

不仅如此,这组方程的不同解答恰与 f 下 $t_1, \cdots, t_s$ 的不同旋向次序相对应。

注释　在 $s=3$ 时,方程组(Q)只有一个方程 $Q_{123}(\varphi) = 1$。这时已知(Q)的解答与 f 下 t_1, t_2, t_3 的旋向次序相对应。在 $s>3$ 时,方程组(Q)的方程个数为 $\binom{s}{3}$,旋数 f_{ij} ($= f_{ji}+1$)的个数为 $\binom{s}{2}$,旋向次序则有 $s!$ 种。因而虽然对每一旋向次序有一确定的方程组的解答,但反之是否对每一组解答有一确定的旋向次序与之相当,并不显然。本术指出答案是正面的,即解答与旋向次序完全对应。为此对 s 用归纳法验证如下。

设 $(\varphi_{ij}) = (c_{ij})$ ($c_{ij} = c_{ji}+1$, i, $j = 1$, $\cdots$, s)是方程组(Q)的一组解答。如前记通道 P_v 上以 v 为端点的树棱为 t_v,又记由 t_v, t_1, $\cdots$, t_k 所构成的星状树形为 T_k,由归

纳假设已知，有嵌入 $f: T_{s-1} \subset$ 平面，使

$$f_{ij} = c_{ij} \ (i, \ j = 1, \ \cdots, \ s-1)。$$

设在 f 下 t_v，t_1，$\cdots$，t_{s-1} 依反钟向旋转时次序为 t_v，t_{i1}，$\cdots$，t_{is-1}，这里（i_1，$\cdots$，i_{s-1}）是（1，$\cdots$，$s-1$）的一个排列。于是有

$$f_{i_m i_n} = 0 \ (m < n, \ m, \ n = 1, \ \cdots, \ s-1)。$$

设 $c_{i_1 s}$，$c_{i_2 s}$，$\cdots$，$c_{i_{s-1} s}$ 中最早一个 $\neq 0$ 的是 $c_{i_{p+1} s}$，即

$$c_{i_1 s} = \cdots = c_{i_p s} = 0,$$

$$c_{i_{p+1} s} = 1。$$

由于对 $q > p+1$ （c_{ij}）满足方程 $Q_{i_{p+1} i_q s}$（φ）$=1$，故有等式

$$c_{i_{p+1} i_q} c_{i_{p+1} s} + c_{i_q s} c_{i_q i_{p+1}} + c_{s i_{p+1}} c_{s i_q} = 1。$$

因

$$c_{i_{p+1} i_q} = f_{i_{p+1} i_q} = 0,$$

$$c_{i_q i_{p+1}} = f_{i_q i_{p+1}} = 1,$$

$$c_{s i_{p+1}} = 0,$$

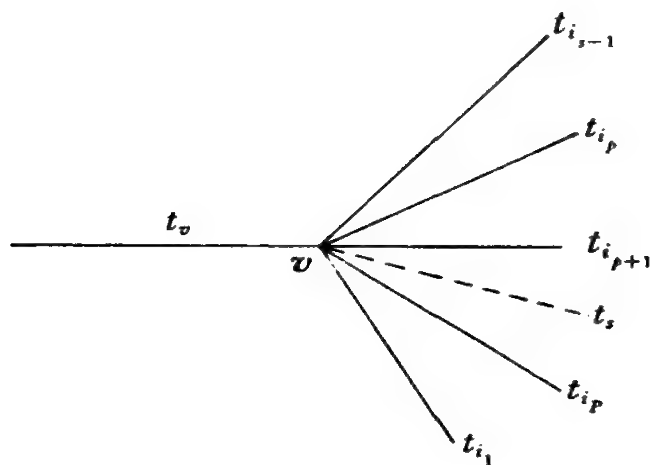


图 11

故得

$$c_{i_q s}=1。$$

今作嵌入 $f: t_s \subset$ 平面，使在 f 下 t_s 位于 t_{i_p} 与 $t_{i_{p+1}}$ 之间，则显然有

$$f_{si}=c_{si}, \quad i=1, \ \cdots, \ s-1,$$

即有嵌入（见图 11）：

$$f: T_s \subset \text{平面},$$

使原来解答 $(\varphi_{ij}) = (c_{ij})$ 与在 f 下反钟向的旋向次序

$$t_v, \ t_{i_1}, \ \cdots, \ t_{i_p}, \ t_s, \ t_{i_{p+1}}, \ \cdots, \ t_{i_{s-1}}$$

相对应。至此验证完毕。

4. 树形嵌入的比较

问题 为对树形 T 的所有嵌入进行分类，首先需对任两嵌入 $f: T \subset$ 平面与 $g: T \subset$ 平面进行比较，试考虑比较的方法。

树形比嵌术 设 $f: T \subset$ 平面，$g: T \subset$ 平面。对以同一顶点 v 为前端的任两树棱 t_i，t_j，置

$$\varepsilon_{ij}=f_{ij}+g_{ij}。$$

则仍以 t_v 表通道 P_v 上以 v 为端点的树棱时，有

$$\varepsilon_{ij}=\begin{cases}0, & \text{在 } f, \ g \text{ 下 } t_v, \ t_i, \ t_j \text{ 旋向相同}, \\ 1, & \text{在 } f, \ g \text{ 下 } t_v, \ t_i, \ t_j \text{ 旋向相反}。\end{cases}$$

显然，数组 (ε_{ij}) 足以比较 f，g 两嵌入的异同：

$$\boxed{f \sim g \Leftrightarrow \text{所有 } \varepsilon_{ij}=0}$$

注意，与 $f_{ij}=f_{ji}+1$，$g_{ij}=g_{ji}+1$ 不同，在诸 ε 间有关系

$$\varepsilon_{ij}=\varepsilon_{ji}。$$

从这些数的几何意义可看出其理由。

5. 树形嵌入的分类

问题 设 T 是树形，则 T 必可嵌入平面。试对不同的嵌

入进行分类。

剖析 由比嵌术，两个不同的嵌入

$$f: T\subset \text{平面}, \quad g: T\subset \text{平面}$$

可由诸数

$$\varepsilon_{ij} = \varepsilon_{ji} = f_{ij} + g_{ij}$$

确定，因而若任取一嵌入 $f: T\subset$平面，则任一其他嵌入将由数组（ε_{ij}）唯一确定。然而诸数 ε_{ij} 不能是任意的，因为依旋数关系，对于任三以同一顶点 v 为前端的树棱 t_i，t_j，t_k，应有

$$Q_{ijk}(f) \equiv f_{ij}f_{ik} + f_{jk}f_{ji} + f_{ki}f_{kj} = 1$$

以及

$$Q_{ijk}(g) \equiv g_{ij}g_{ik} + g_{jk}g_{ji} + g_{ki}g_{kj} = 1。$$

后者也可写作

$$(f_{ij} + \varepsilon_{ij})(f_{ik} + \varepsilon_{ik}) + (f_{jk} + \varepsilon_{jk})(f_{ji} + \varepsilon_{ji}) + (f_{ki} + \varepsilon_{ki})(f_{kj} + \varepsilon_{kj}) = 1。$$

换言之，若引入一些（模 2）未知数

$$x_{ij} = x_{ji},$$

则（ε_{ij}）应是下面一组方程的解：

$$(f_{ij} + x_{ij})(f_{ik} + x_{ik}) + (f_{jk} + x_{jk})(f_{ji} + x_{ji}) + (f_{ki} + x_{ki})(f_{kj} + x_{kj}) = 1。$$

反之，若$(x_{ij}) = (\varepsilon_{ij})$是上组方程的一个解答，则

$$(\varphi_{ij}) = (f_{ij} + \varepsilon_{ij})$$

满足方程组

$$(Q) \qquad \varphi_{ij}\varphi_{ik} + \varphi_{jk}\varphi_{ji} + \varphi_{ki}\varphi_{kj} = 1。$$

因而由一般旋数术，对于任一顶点 v，有一嵌入使通道 P_v 上的 t_v 以及以 v 为前端的诸树棱在此嵌入下有一确定旋向，使相应旋数即为上述诸数 $f_{ij} + \varepsilon_{ij}$。今作嵌入 $g: T\subset$平面，使在

g 下每一顶点处诸棱的旋向都恰如上述，则显然有

$$g_{ij} = f_{ij} + \varepsilon_{ij}。$$

由此得到下述结论。

嵌树术 设 $f: T \subset$ 平面是树形 T 的一个固定嵌入，则 T 在平面中的所有不同嵌入与下述方程组的解答（$x_{ij} = x_{ji}$）

$$(x_{ij}) = (\varepsilon_{ij})$$

一一对应：

$$\boxed{Q_{ijk}(f, x) = 1}$$

其中

$$\begin{aligned} Q_{ijk}(f, x) = &(f_{ij} + x_{ij})(f_{ik} + x_{ik}) \\ &+ (f_{jk} + x_{jk})(f_{ji} + x_{ji}) \\ &+ (f_{ki} + x_{ki})(f_{kj} + x_{kj})。\end{aligned}$$

诸方程则与以同一顶点为前端的三棱组 $\langle t_i, t_j, t_k \rangle$ 相对应。

三、线图的嵌入问题

1. 交截数

举例 已知任一树形必可嵌入平面，但若线图中有闭圈(或称回路)，嵌入就不一定可能。或者即使可能，若将线图中所含树形嵌入后，是否可将嵌入扩充至所余诸棱，也当视树形的嵌入是否适当而定。

例如设线图 G 有 5 个顶点 v_0，v_1，…，v_4 和 6 个棱 t_1，…，t_4，e_{12}，e_{34}，其中 t_i 连结 v_0 与 v_i（$i = 1$，…，4），e_{12} 与 e_{34} 各连结 v_1，v_2 与 v_3，v_4。命 t_1，…，t_4 所成树形为 T。若嵌入 T 在反钟向旋转时依次为 t_1，t_2，t_3，t_4，则可有三种不同的方式扩充这一嵌入至整个线图 G，见附图 12。但

若嵌入 T 在反钟向旋转时依次为 t_1，t_3，t_2，t_4，如附图 13，即不能将这一嵌入扩充至 G，原因甚为显然：若先扩充 e_{12} 的嵌入，则不论嵌入成图 14 或图 15，e_{12} 与 t_1，t_2 所围成的闭圈将把平面分成内外两部分，v_3 与 v_4 必一在闭圈之内而另一在闭圈之外。因而将连结 v_3，v_4 的 e_{34} 嵌入平面时必将与闭圈相交，而不能得一扩充至整个线图 G 的嵌入。

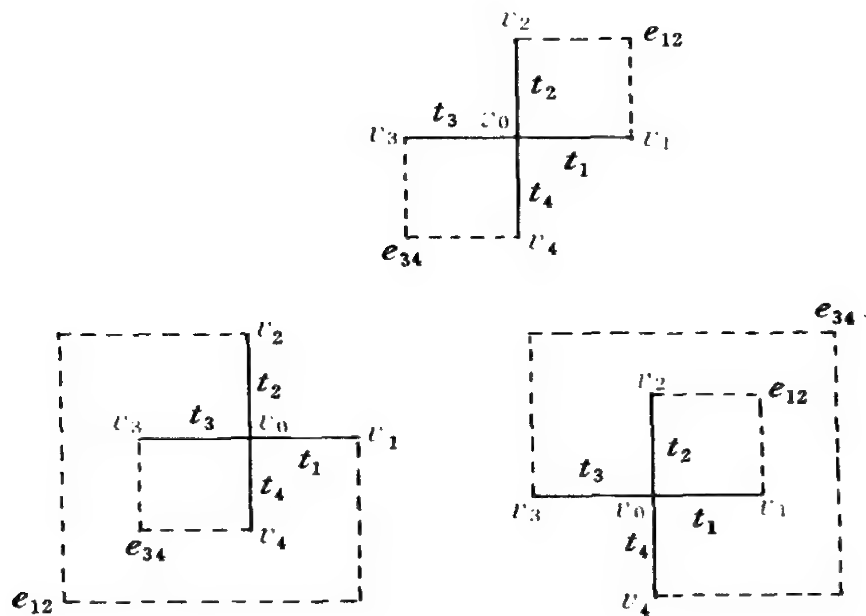


图 12

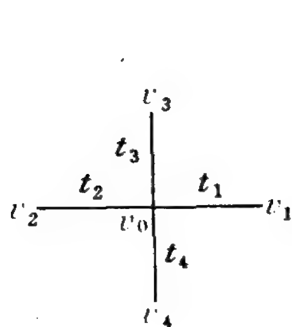


图 13

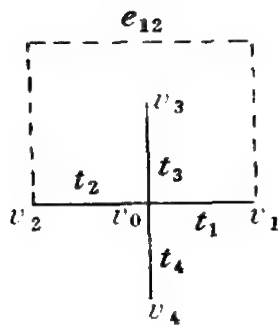


图 14

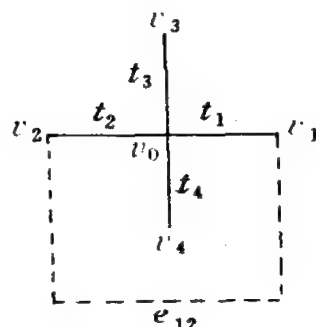


图 15

上面提到了这样一个事实：平面上的任一闭圈把平面分成内外两部分。同在内部（或同在外部）的任意两点都可用完全处于内部（或外部）的折线相连结。但若两点一在内部而另一在外部，则任一连线必与圈线相遇。

这一事实国外通称 Jordan **定理**。我们认为可迳直称之为 Jordan **公理**。作为定理来加以证明，不是绝对必要的。事实上，在平面拓扑学中这也是作为平面公理系统的公理之一来引入的。即使在 Hilbert 的初等几何学公理系统中，也有好几条颇为繁琐的平面次序公理，其作用犹如这一 Jordan 公理。

以下将上述直观的经验事实称为**闭圈原理**。它是线图有时不能嵌入平面的根本原因，也是布线之所以成为较难的问题的根本原因。

以后，我们还需要下面较一般的形式，例见图 16。

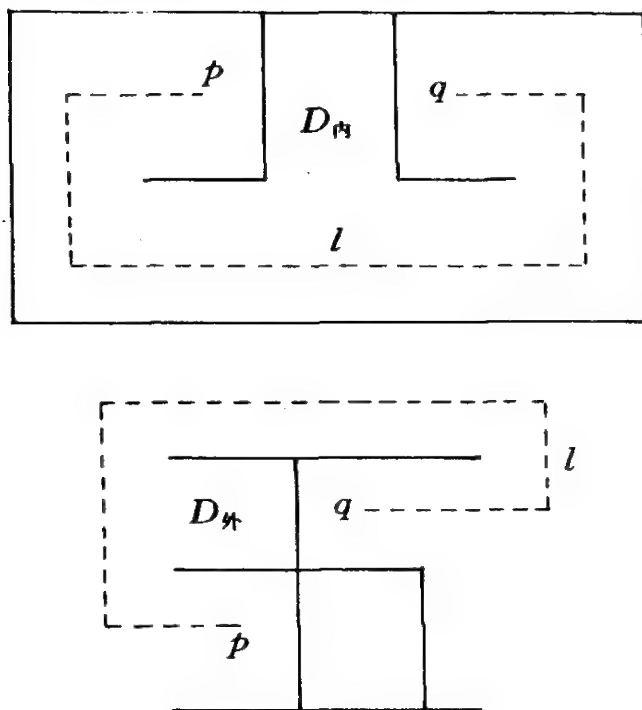


图 16

设平面上有一闭圈 C 以及 C 内部（与外部）若干通向 C 而互不相遇的折线，则平面上不在 C 与这些折线上的点构成两个**区域** $D_{内}$ 与 $D_{外}$，各以 C 以及这些分别在 C 内与 C 外的折线为其**边界集**，记作 $\dot{D}_{内}$ 与 $\dot{D}_{外}$。同在 $D_{内}$ 与 $\dot{D}_{内}$ 或同在 $D_{外}$ 与 $\dot{D}_{外}$ 的任意两点 p，q，可用一折线 l 在 C 内部或外部相连，而这一折线 l 至多除 p，q 外都不与 $\dot{D}_{内}$ 和 $\dot{D}_{外}$ 相遇。但若 p，q 分处 $D_{内}$ 与 $D_{外}$，则任一连结 p，q 的折线必与 C 相遇。

此外，我们将引入**交截数**的概念，使闭圈原理有定量的描述。

交截术　平面上一圈线 C 分平面成内外两部分，若折线 l 的两端 p，q 都不在 C 上，而 l 与 C 在每一可能有的交点处相截，则 p，q 同在 C 内部或同在 C 外部时，l 与 C 的交截点数必为偶数。反之，若 p，q 一在 C 内而另一在 C 外时，l 与 C 的交截点数必为奇数。这一基本事实也可用公式表达如下。

引入 l 与 C 的**模 2 交截数**（简称**交截数**）：

$$交（l，C）=\begin{cases}0, & l\text{ 与 }C\text{ 交截点数为偶数时，}\\ 1, & l\text{ 与 }C\text{ 交截点数为奇数时。}\end{cases}$$

又引入模 2 数：

$$次_pC=\begin{cases}0, & p\text{ 在 }C\text{ 外部时，}\\ 1, & p\text{ 在 }C\text{ 内部时，}\end{cases}$$

则

$$交（l，C）=次_pC+次_qC。$$

又平面中只在可能有的交点处相截的任两线图 l_1，l_2 也可定一（模 2）**交截数**，记作交（l_1，l_2）。且这些交截数是**加法性**的，例如

$$交（l_1，l_2\cup l_3）=交（l_1，l_2）+交（l_1，l_3）。$$

2. 方法概述

G 是一遵守第一部分第 2 节中那些限制的线图，需要考虑 G 在平面中能否嵌入的问题。在 G 无闭圈即为一树形时，嵌入问题已由第二部分完全解决。前节例中已指出，对于一般的线图 G，嵌入问题的困难乃是由闭圈所引起。这些考虑导致我们采用下述方法。

在 G 中选择一个含有 G 中所有顶点的树形，称为 G 的**最大树**，记为 T。相对于一固定的 T 来说，T 中的棱称为**树棱**，记为 t_i，t_j 等。G 中不属于 T 的棱将称为**外棱**，记作 e_α，e_β 等。若外棱 e_α 的两端为 a，a'，则在 T 中有一唯一的通道连结 a 与 a'，记作 P_α。闭圈 $e_\alpha \cup P_\alpha$ 记作 C_α。

我们的方法可概述如下。

先将 T 任意嵌入平面，记嵌入为 $f: T \subset$ 平面。然后考虑将 f 扩充至诸外棱 e_α。一般说来，这是不可能的。我们将引入一（模 2）数组来定量描述这一不可能的程度。这种不可能或者是由于原来嵌入 f 的选择不适当所引起（见前举例），或者由于线图 G 的内在矛盾所引起。我们将引入另一（模 2）数组来定量刻划这些内在矛盾，这两数组间关系的确定，导致能否改变原来的嵌入 $f: T \subset$ 平面，使扩充至整个 G 的嵌入成为可能。

在描述树形 T 的嵌入时，由二知，若事先选一树根可有许多方便。又由于在尝试将 $f: T \subset$ 平面扩充至诸外棱 e_α，e_β 等时，需要考虑 e_α，e_β 嵌入平面后的交截数，若 e_α，e_β 有公共端点，则嵌入时交截数的叙述较为繁琐。为了避免出现两个外棱有公共顶点这种情形，以及为了树根的引入，特先将线图 G 与最大树 T 加以适当改造。先举两例如下（实线为最大树，虚线为外棱）：

例 1　Kuratowski 线图 K_1。

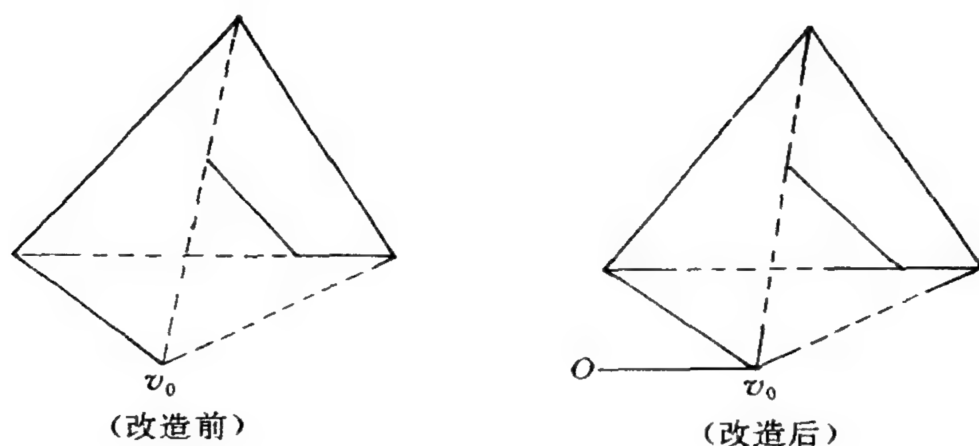


图 17

例 2　Kuratowsiki 线图 K_2。

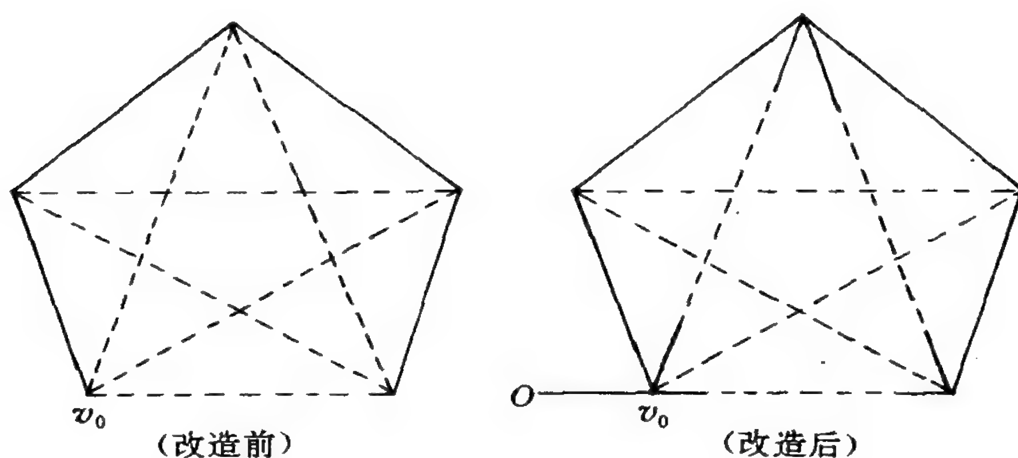


图 18

总的说来，线图的**改造**包括两个方面：

(1) 将任一外棱 e_a 分成三段，在 e_a 两端的两段归入最大树 T，中间的一段则作为新的外棱。

(2) 在 T 中任取一点 v_0（不必为顶点），并添入一顶点 O 作为**树根**以及自根 O 至 v_0 的一个树棱。

显然，第一种改造对于线图的嵌入无丝毫影响。对于第二

种改造，如果 v_0 是一个 k **叉点**，即在 v_0 处恰有 k 个棱，对原线图的每一嵌入，改造后的线图恰有 k 个不同嵌入与之相当，因而解决了改造后线图的嵌入问题，同时也就解决了原来线图的嵌入问题。

由于上述，我们就已经改造的线图（仍记之为 G）与最大树（仍记之为 T）进行讨论，引入的树根记作 O，这时 G，T 等具有以下诸性质：

1. 自最大树 T 的任一自由端出发，至多只有一个外棱。

2. 任两外棱均无公共端点。

3. 在树根 O 处无外棱。

以下假定 G，T 与根 O 已固定，不再声明。

3. 矛盾数

问题 设 $f: T\subset$ 平面。一般说来，f 不能扩充为整个线图 G 的嵌入 $f: G\subset$ 平面，使 $f|T\equiv f$。问如何刻划这种不可能的程度。

扩充术 设任意扩充 f 为相对于 T 的浸入 $f: G \underset{T}{\propto}$ 平面，使 $f|T\equiv f$。对 T 外的任两外棱 e_α，e_β，

$$\text{交}\ (f\ (e_\alpha),\ f\ (e_\beta))$$

与扩充方式 f 无关，而由 $f|T$ 即 f 完全确定，记之为 $f_{\alpha\beta}$。于是

$f: T\subset$ 平面可扩充为 $f: G\subset$ 平面，使 $f|T\equiv f\Leftrightarrow$ 一切 $f_{\alpha\beta}=0$

由此知，数组（$f_{\alpha\beta}$）可作为能否扩充 $f: T\subset$ 平面为整个 G 的嵌入的一种量度。换言之，使 f 不能扩充为 G 的嵌入的内部矛盾暴露于外时，可通过数组（$f_{\alpha\beta}$）表现出来。为此，$f_{\alpha\beta}$ 将称为相应于 $f: T\subset$ 平面的**矛盾数**，（$f_{\alpha\beta}$）称为**矛盾数**

组。

注析 设 $f':G_T\subset$ 平面是 $f:T\subset$ 平面的另一扩充。不妨设 $f'(e_\alpha)$ 与 $f(e_\beta)$ 相截（一般情形容易归结于此），因树 T 与 (e_α) 在 f 与 f' 下都不相遇，故

$$交(f(e_\alpha),f(e_\beta))=交(f(e_\alpha),f(C_\beta)),$$
$$交(f'(e_\alpha),f(e_\beta))=交(f'(e_\alpha),f(C_\beta))。$$

又折线 $f(e_\alpha)$，$f'(e_\alpha)$ 有相同端点，故由第三部分第一节的交截术知，右端两式相等。由此得

$$交(f(e_\alpha),f(e_\beta))=交(f'(e_\alpha),f(e_\beta))。$$

同样有

$$交(f'(e_\alpha),f'(e_\beta))=交(f'(e_\alpha),f(e_\beta)),$$

故

$$交(f'(e_\alpha),f'(e_\beta))=交(f(e_\alpha),f(e_\beta)),$$

即交 $(f(e_\alpha),f(e_\beta))$ 与扩充方式号无关。这说明了符号 $f_{\alpha\beta}$ 的合理性。

显然，在 $f:T\subset$ 平面可扩充为整个 G 的嵌入时应有一切 $f_{\alpha\beta}=0$。反之也容易由归纳法验证如下：

将一切外棱任意排列成一次序 $e_1,\cdots,e_L$，将树添入 $e_1,\cdots,e_k$（对 $k=1,\cdots,L$）后所得线图记作 G_k。又置 $G_0=T$。设 $f:T\subset$ 平面已扩充为一嵌入（$f_0\equiv f:G_0\subset$ 平面）：

$$f_k:G_k\subset 平面,$$
$$f_k\mid T\equiv f,$$

试考虑继续扩充到 G_{k+1} 的可能性。$f(G_k)$ 分割平面成若干区域，设为 $D_1,\cdots,D_r$。设 e_{k+1} 的两端为 a_{k+1},a'_{k+1}，而 $f_k(a_{k+1})=p_{k+1}$，$f_k(a'_{k+1})=p'_{k+1}$。将 p_{k+1}，p'_{k+1} 在平面中任以一折线 l 相连，而使 l 除 p_{k+1}，p'_{k+1} 外不与 $f_k(T)$ 相遇且与每一 $f_k(e_i)$ 至多相截。每一区域 D_i 的边界集 $\dot{D}_i$ 由一闭圈 C_i 以及若

干在 C_i 内通向 C_i 而互不相遇的折线所构成，而 C_i 又由 $f_k(T)$ 的一部分以及若干 $f_k(e_j)(1\leqslant j\leqslant k)$ 所组成，设为 $f_k(e_{ij})$，$j=1,\cdots,s(i)$，折线则全由树棱所组成。

由假设知(下面符号指 $f_{\alpha\beta}$，应不致混淆为 f_{ij})：

$$f_{k+1,1}=\cdots=f_{k+1,k}=0,$$

因而

$$交(l,C_i)=\sum_{j=1}^{s(i)}交(l,f_k(e_{ij}))=\sum_{j=1}^{s(i)}f_{k+1},i_j=0。$$

由此据一般形式的闭圈原理知：若 p_{k+1} 在某一区域 D_i 的边界且在闭圈 C_i 的内部，则 p'_{k+1} 也将同在 C_i 的内部，因而在 C_i 内部可将 p_{k+1}，p'_{k+1} 用一折线 l_{k+1} 连结，使 l_{k+1} 与树形 $f_k(T)$ 除 p_{k+1}，p'_{k+1} 外不相遇，且 l_{k+1} 不与任一 $f_k(e_1),\cdots,f_k(e_k)$ 相遇，故可将 f_k 扩充至 e_{k+1}，得 $f_{k+1}:G_{k+1}\subset$平面，使

$$f_{k+1}(e_{k+1})=l_{k+1},$$

$$f_{k+1}|G_k\equiv f_k。$$

依次进行即可将 f 扩充为整个 G 的嵌入

$$f:G\subset平面,$$

$$f/T\equiv f。$$

验证完毕。

4. 基本关系式

问题 设 $f:T\subset$平面，则 f 能否扩充为嵌入 $f:G\subset$平面，使 $f|T\equiv f$ 的可能程度由矛盾数组（$f_{\alpha\beta}$）刻划。矛盾的产生导源于 T 的嵌入 f，而这一嵌入可由数组（f_{ij}）该划（见前二）。在外部表现（$f_{\alpha\beta}$）与内在原因（f_{ij}）之间必有关系，试求这些关系的明确表达式。

基本关系术 对于 T 的任两自由端 a 与 b，命 T 中连结 a，b 与根 O 的通道 P_a，P_b 首次相遇的顶点为 v。命在 T 中

连结 v 与 a，b 的通道 P_{va}，P_{vb} 上以 v 为前端的树棱为 t_i，t_j。引入记号

$$f_{ab}=f_{ij},$$

显然有

$$f_{ab}=f_{ba}+1。$$

例如在附图 19 中有

$$f_{ab}=f_{ij}=0。$$

设外棱 e_α 的两端为 a，a'：$e_\alpha=\widehat{aa'}$，

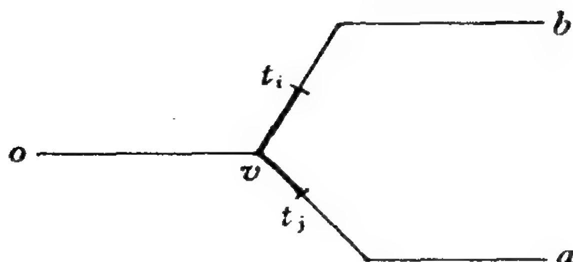


图 19

外棱 e_β 的两端为 b，b'：$e_\beta=\widehat{bb'}$，则有**基本关系式**

$$\boxed{f_{\alpha\beta}=f_{ab}+f_{ab'}+f_{a'b}+f_{a'b'}}$$

式中右边与 a，a'的次序无关，也与 b，b'的次序无关。两边都与 α，β 的次序无关。

注析　任意扩充 $f: T\subset$ 平面为 $f: G\underset{T}{\propto}\subset$ 平面，就 f 下 O 在圈线 $C_\beta=e_\beta\cup P_\beta$ 的内外以及 P_α 与 C_β 的相对位置关系可有以下几种如图 21 所示的不同情形，其它情形可由交换 b，b' 而得。

引入记号

$$f_{a\beta}=f_{ab}+f_{ab'},$$

则依据这些情形可列成下表：

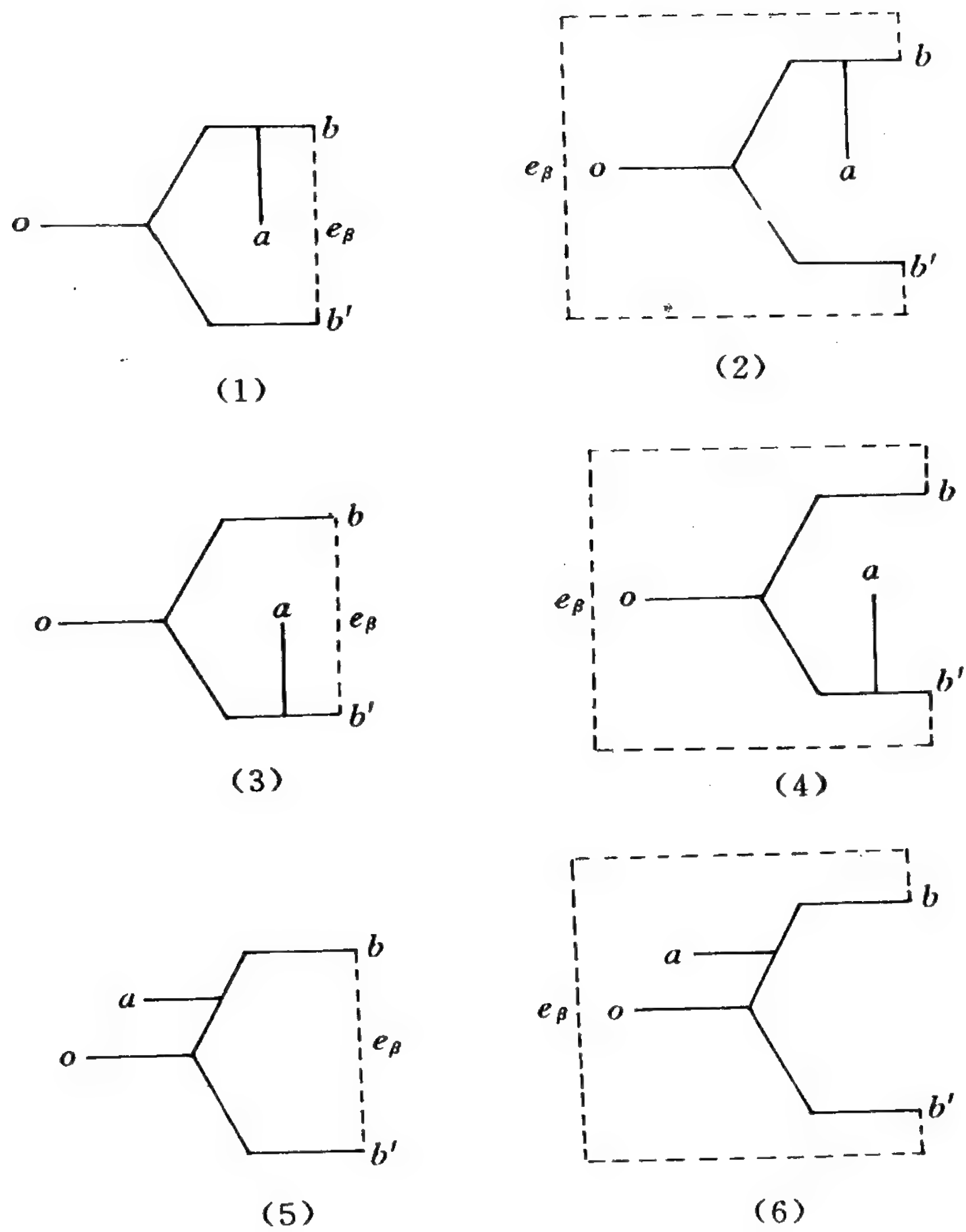
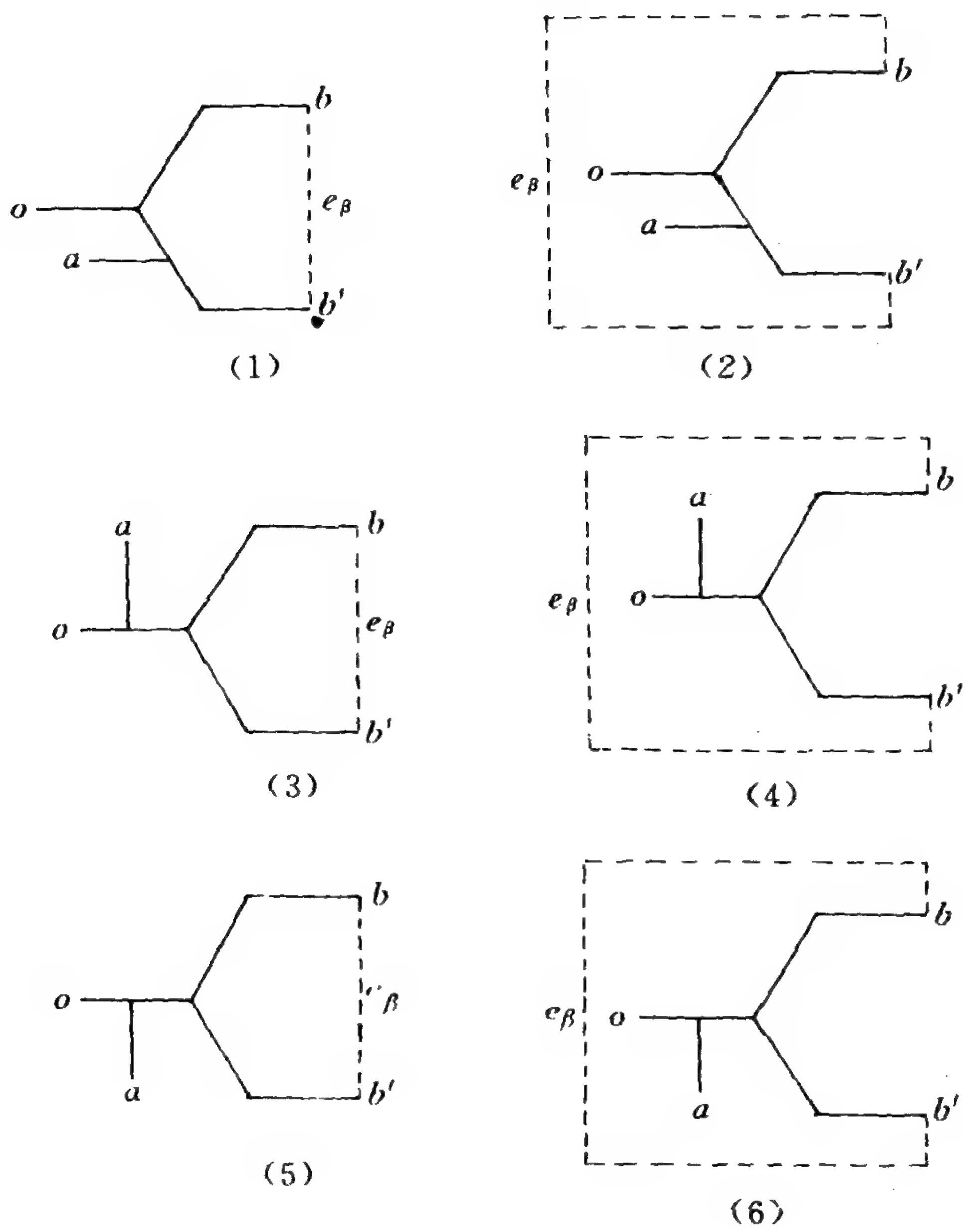


图 20



b
e_β
o
a
b'
(1)
b
e_β
o
a
b'
(2)
a
b
o
e_β
b'
(3)
b
a
e_β
o
b'
(4)
b
o
b'
a
(5)
b
e_β
o
a
b'
(6)

图 21

情　　形	f_{ab}	$f_{ab'}$	$f_{\alpha\beta}$	次 $a\,\bar{f}(C_\beta)$	次 $a\,\bar{f}(C_\beta)$
1	0	1	1	1	0
2	0	1	1	0	1
3	0	1	1	1	0
4	0	1	1	0	1
5	1	1	0	0	0
6	1	1	0	1	1
7	0	0	0	0	0
8	0	0	0	1	1
9	1	1	0	0	0
10	1	1	0	1	1
11	0	0	0	0	0
12	0	0	0	1	1

上表中 $f_{bb'}=1$。若 $f_{bb'}=0$，则只需将 f_{ab}，$f_{ab'}$ 的值互易即可，此外勿须变更。

由表可知，不论何时都有

$$f_{a\beta}=次_{a}f(C_\beta)+次_{O}f(C_\beta)。$$

同样有

$$f_{a'\beta}=次_{a'}f(C_\beta)+次_{O}f(C_\beta)。$$

相加得

$$f_{a\beta}+f_{a'\beta}=次_{a}f(C_\beta)+次_{a'}f(C_\beta)。$$

由第1节的交截术知，

$$上式=交(f(e_\alpha),f(C_\beta))=交(f(e_\alpha),f(e_\beta))=f_{\alpha\beta}。$$

故

$$f_{\alpha\beta}=f_{a\beta}+f_{a'\beta}=f_{ab}+f_{ab'}+f_{a'b}+f_{a'b'}。$$

此即基本关系式。

5. 线图嵌入第一基本定理

问题　试从 $f: T\subset$平面所定数组(f_{ij})求 f 可改变为 $g: T\subset$平面，而 g 可扩充为整个 G 的嵌入的充要条件。

剖析　任两嵌入 $f, g: T\subset$平面的异同可由一数组(ε_{ij})确定（见Ⅱ中的树形比嵌术与嵌树术），此处

$$\varepsilon_{ij}=\varepsilon_{ji}=f_{ij}+g_{ij}。$$

又对 $g: T\subset$平面由第3节扩充术已知 g 可扩充为整个 G 的嵌入的充要条件。由此易得 $f: T\subset$平面可改变为 $g: T\subset$平面，而 g 可扩充为整个 G 的嵌入的充要条件。

详言之，设 $f, g: T\subset$平面任意，对 T 的任两自由端 a, b，命 a, b 到根 O 的通道 p_a, P_b 最初相遇的顶点为 v。连结 v 与 a, b 的通道 P_{va}, P_{vb} 上以 v 为前端的树棱设为 t_i, t_j，则置

$$\varepsilon_{ab}=\varepsilon_{ij}。$$

又对任两外棱 $e_\alpha=\widehat{aa'}, e\beta=\widehat{bb'}$，置

$$\varepsilon_{\alpha\beta}=\varepsilon_{ab}+\varepsilon_{ab'},$$

$$\varepsilon_{ab}=\varepsilon_{ab}+\varepsilon_{a'b},$$

$$\varepsilon_{\alpha\beta}=\varepsilon_{ab}+\varepsilon_{ab'}+\varepsilon_{a'b}+\varepsilon_{a'b'},$$

则显有

$$\varepsilon_{ab}=\varepsilon_{ba},$$

$$\varepsilon_{\alpha\beta}=\varepsilon_{\beta\alpha},$$

$$\varepsilon_{ab}=\varepsilon_{ba},$$

$$\varepsilon_{\alpha\beta}=\varepsilon_{\beta\alpha}=\varepsilon_{\alpha\beta}+\varepsilon_{\alpha'\beta}=\varepsilon_{ab}+\varepsilon_{ab'}。$$

由

$$g_{ij}=f_{ij}+\varepsilon_{ij},$$

依据诸符号的意义以及上一节的基本关系术可得

$$g_{ab} = f_{ab} + \varepsilon_{ab},$$

$$g_{\alpha\beta} = f_{\alpha\beta} + \varepsilon_{\alpha\beta}。$$

因而由扩充术，若 f 可改变为 $g: T\subset$平面而 g 可扩充为整个 G 的嵌入，则 $g_{\alpha\beta}=0$，或即

$$\varepsilon_{\alpha\beta} = f_{\alpha\beta}。$$

今依第二部第5节引入一组(模2)未知数 $x_{ij} = x_{ji}$，又对任两 T 的自由端 a, b 与任两外棱 $e_\alpha = \widehat{aa'}$，$e_\beta = \widehat{bb'}$，置(i, j 意义同前)

$$x_{ab} = x_{ij},$$

$$x_{\alpha\beta} = x_{ab} + x_{ab'},$$

$$x_{ab} = x_{ab} + x_{a'b},$$

$$x_{\alpha\beta} = x_{ab} + x_{ab'} + x_{a'b} + x_{a'b'},$$

则显然有

$$x_{ab} = x_{ba},$$

$$x_{\alpha\beta} = x_{\beta\alpha},$$

$$x_{ab} = x_{ba},$$

$$x_{\alpha\beta} = x_{\beta\alpha} = x_{\alpha\beta} + x_{\alpha'\beta} = x_{ab} + x_{ab'}。$$

于是上述 $g: T\subset$平面可扩充为整个 G 的嵌入的条件可表为：(ε_{ij})应为下方程组

(Ⅰ)$_j$ $$x_{\alpha\beta} = f_{\alpha\beta}$$

的解。又据Ⅱ5嵌树术，(ε_{ij})还应满足以下方程组：

(Ⅱ)$_j$ $$Q_{ijk}(f, x) = 1,$$

其中 $Q_{ijk}(f, x)$为一与以同一顶点为前端的三棱组 t_i, t_j, t_k 相对应的二次式。

反之，若(ε_{ij})满足这两方程组，则依(Ⅱ)$_f$ 与嵌树术可据 ε_{ij} 以改变 f 为一嵌入 $g: T\subset$平面，使

$$g_{ij} = f_{ij} + \varepsilon_{ij},$$

其次依(Ⅰ)$_f$ 有

$$q_{\alpha\beta}=0。$$

因而由扩充术可将 g 扩充为一整个 G 的嵌入。

综合上述可得出问题的完全解答如下:

线图嵌入术(第一基本定理) 对于以任意同一顶点为前端的任两树棱 t_i, t_j 引入一模 2 未知数 $x_{ij}=x_{ji}$,又对 T 的任两自由端 a, b 与任两外棱 e_α, e_β 引入

$$x_{ab}=x_{ij},$$

$$x_{\alpha\beta}=x_{ab}+x_{ab'}+x_{a'b}+x_{a'b'},$$

其中 t_i, t_j 为自 a, b 至 O 的通道 P_a, P_b 上以初次相遇的顶点为前端的两树棱,而 $e_\alpha=\widehat{aa'}, e_\beta=\widehat{bb'}$。对每一对外棱 e_α, e_β 引入方程

(Ⅰ)$_f$ $$x_{\alpha\beta}=f_{\alpha\beta},$$

又对每三以同一顶点为前端的树棱 t_i, t_j, t_k 引入方程

(Ⅱ)$_f$ $$Q_{ijk}(f,x)=1,$$

此处

$$\begin{aligned}Q_{ijk}(f,x)=&(f_{ij}+x_{ij})(f_{ik}+x_{ik})\\&+(f_{jk}+x_{jk})(f_{ji}+x_{ji})\\&+(f_{ki}+x_{ki})(f_{kj}+x_{kj}),\end{aligned}$$

则 $f: T\subset$平面可改变为 $g: T\subset$平面而 g 可扩充为整个 G 的嵌入的充要条件为方程组(Ⅰ)$_f$,(Ⅱ)$_f$ 有解。

简言之,有

G 能嵌入平面$\Leftrightarrow$(Ⅰ)$_f$,(Ⅱ)$_f$ 有解

这一结论将称为**线图嵌入第一基本定理**。

若已得(Ⅰ)$_f$,(Ⅱ)$_f$ 的一组解答$(x_{ij})=(\epsilon_{ij})$,则视 $\epsilon_{ij}=1$ 或 0 而确定是否改变 t_i, t_j 的旋向,即可从 f 得一确定的嵌入g:

$T\subset$平面而 g 可扩充为整个 G 的嵌入，如扩充术所示。

6. 不能嵌入平面的线图实例

问题 举出一些不能在平面中嵌入的简单线图。

Kuratowski 实例。Kuratowski（波兰数学家）在 1930 年曾举出过两个这样的线图，记作 K_1, K_2，见图 17，18。我们将依据前面线图嵌入术对 K_1, K_2 进行分析并证实 K_1 与 K_2 不能嵌入平面。

先考虑 K_1。为此选一最大树 T 并依第 2 节改造 K_1 且作 $f: K_1 \underset{T}{\subset}$ 平面如图 22（参阅图 17）

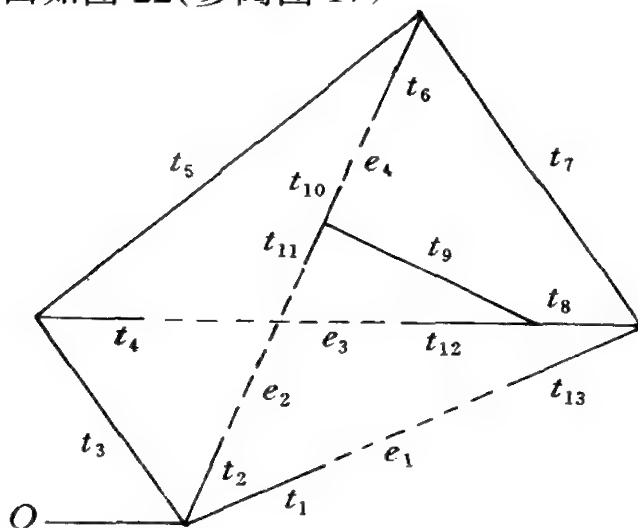


图 22

相当于方程组$(\text{I})_f$，有：

$$(e_1, e_2): x_{8,13} + x_{23} + x_{13} + x_{12} = 0,$$

$$(e_1, e_3): x_{8,13} + x_{45} = 0,$$

$$(e_2, e_3): x_{9,12} + x_{45} = 1,$$

$$(e_1, e_4): x_{8,13} + x_{67} = 0,$$

$$(e_2, e_4): x_{10,11} + x_{67} = 0,$$

$$(e_3, e_4): x_{9,12} + x_{67} = 0。$$

易见上述方程组无解，因而 K_1 不能嵌入平面。

但若弃去 e_4(因而也同时弃去 t_6, t_{10}),则对余下的线图 K'_1(见图 23)在 f 下的相应方程组(Ⅰ)$_f$ 只含有前面三个,这时(Ⅰ)$_f$ 显然可解。因(Ⅱ)$_f$ 并不存在,故知 K'_1 必可嵌入平面。

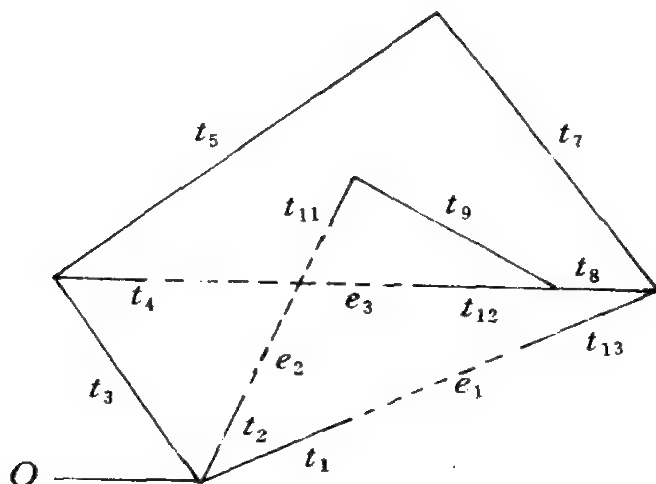


图 23

为求得 K'_1 的不同嵌入,可先求得(Ⅰ)$_f$ 的一般解答如下:

$$x_{23} + x_{13} + x_{12} = x_{8,13} = x_{45},$$

$$x_{9,12} = x_{45} + 1.$$

无损于一般性,不妨设

$$x_{12} = x_{13} = 0,$$

于是有两组不同解答如下:

(1) $x_{12} = x_{13} = x_{23} = x_{45} = x_{8,13} = 0,\ x_{9,12} = 1,$

(2) $x_{12} = x_{13} = x_{9,12} = 0,\ x_{23} = x_{45} = x_{8,13} = 1$。

相应的树形嵌入见图 24,25。

每一树形的嵌入都可扩充为 K'_1 的嵌入,例如在情形(1),可扩充成 4 种不同的嵌入如图 26 所示(树根已省去)。

同样,在情形(2),也可得出 4 种不同的扩充。与图 26 的 4 种扩充一起共 8 种,概括了 K'_1(即 K 除去 e_4 后所余线图)在平面中一切可能的不同嵌入。

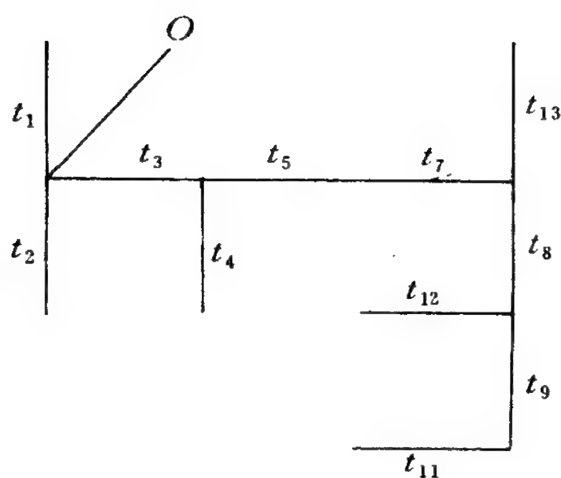


图 24

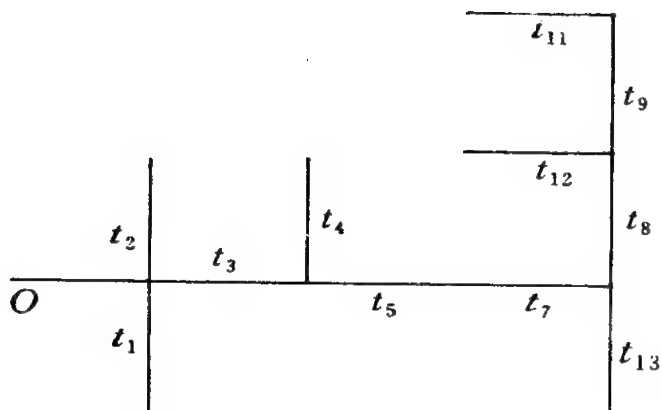
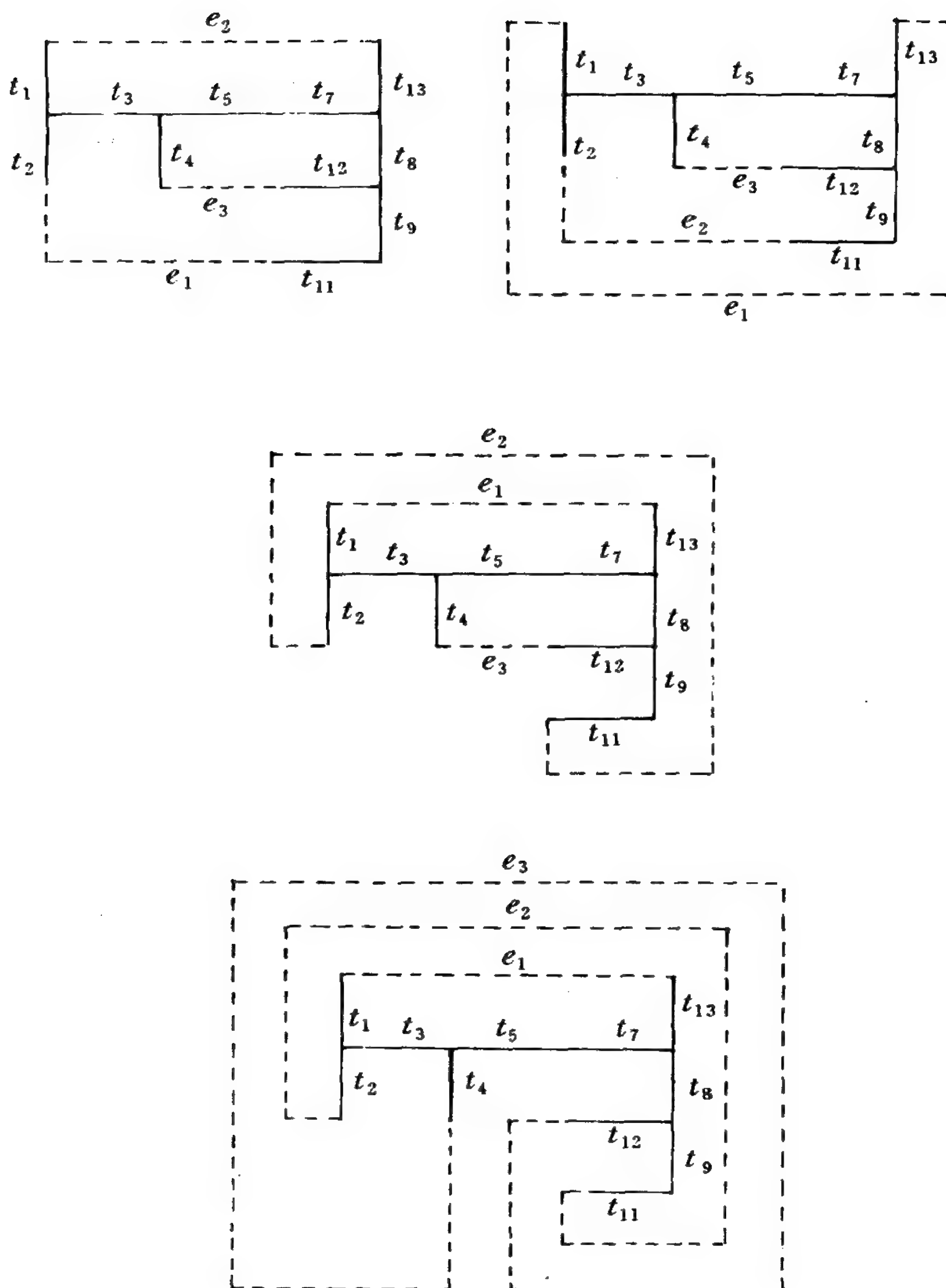


图 25

其次考虑第二种 Kuratowski 线图 K_2(参阅图 18)。为此,取树形 T 并改造 K_2 且作 $f: K_2 \underset{T}{\subset\!\!\!\circ}$ 平面如图 27。

方程组$(\mathrm{I})_f$ 含有 15 个方程如下:

$$(e_1, e_2): \quad x_{12} + x_{14} + x_{24} + x_{12,13} = 0,$$
$$(e_1, e_3): \quad x_{13} + x_{14} + x_{34} + x_{8,10} = 0,$$
$$(e_2, e_3): \quad x_{23} + x_{24} + x_{34} + x_{8,10} = 0,$$
$$(e_1, e_4): \quad x_{57} + x_{15,16} = 0,$$
$$(e_2, e_4): \quad x_{57} + x_{12,13} = 1,$$
$$(e_3, e_4): \quad x_{57} + x_{8,10} = 1,$$



e_2
t_1 t_3 t_5 t_7 t_{13}
t_2 t_4 t_{12} t_8
e_3 t_9
e_1 t_{11}
t_1 t_3 t_5 t_7 t_{13}
t_2 t_4 t_8
e_3 t_{12} t_9
e_2 t_{11}
e_1
e_2
e_1
t_1 t_3 t_5 t_7 t_{13}
t_2 t_4 t_8
e_3 t_{12} t_9
t_{11}
e_3
e_2
e_1
t_1 t_3 t_5 t_7 t_{13}
t_2 t_4 t_8
t_{12} t_9
t_{11}

图 26

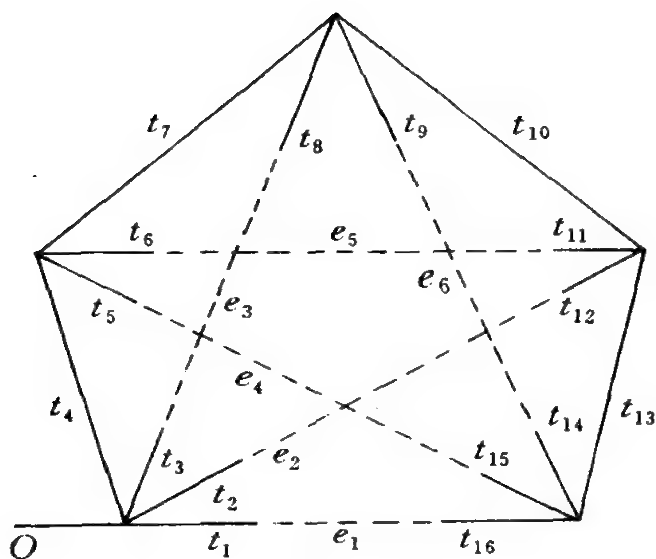


图 27

$$
\begin{aligned}
&(e_1, e_5): && x_{67} + x_{11,13} = 0, \\
&(e_2, e_5): && x_{67} + x_{11,12} = 0, \\
&(e_3, e_5): && x_{67} + x_{8,10} = 1, \\
&(e_4, e_5): && x_{56} + x_{57} + x_{67} + x_{11,13} = 0, \\
&(e_1, e_6): && x_{9,10} + x_{14,16} = 0, \\
&(e_2, e_6): && x_{9,10} + x_{12,13} = 1, \\
&(e_3, e_6): && x_{89} + x_{8,10} = 0, \\
&(e_4, e_6): && x_{9,10} + x_{14,15} = 0, \\
&(e_5, e_6): && x_{9,10} + x_{11,13} = 1。
\end{aligned}
$$

易见上述方程组无解，因而 K_2 不能嵌入平面。但若弃去 e_6(因而也弃去 t_9, t_{14})，即易 K_2 为图 28,29 的 K'_2，并取最大树 T' 与浸入 $f: K'_2 \underset{T'}{\subset}$ 平面如前，则相应的方程组(Ⅰ)$_f$ 系由前面 10 个所组成。易见此时(Ⅰ)$_f$ 有解。

无损于一般性，可设

$$x_{12} = x_{13} = x_{14} = 0,$$

此时方程组(Ⅰ)$_f$ 的一般解答为

$$x_{23}=x_{24}=x_{34}=x_{8,10}=x_{12,13},$$

$$x_{56}=x_{57}=x_{67}=x_{11,12}=x_{11,13}=x_{15,16},$$

$$x_{56}=x_{23}+1。$$

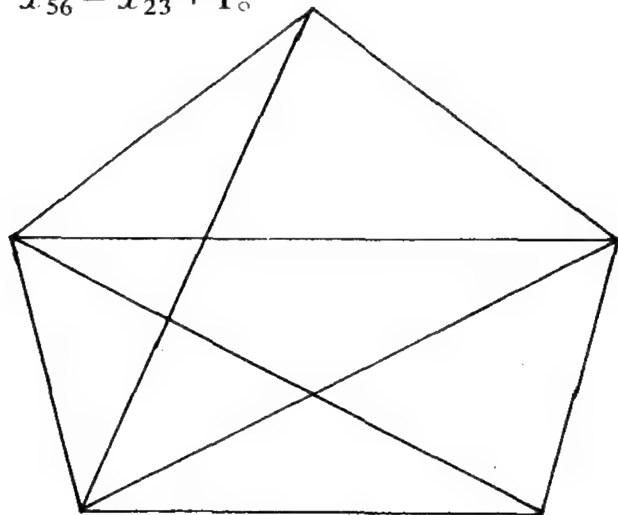


图 28

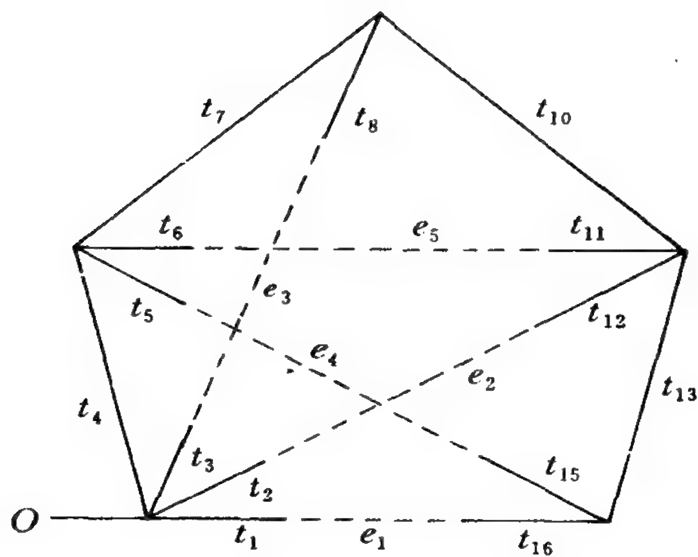


图 29

对于方程组$(\mathrm{II})_f$，此时有相应于

$$(i,j,k)=(123),(124),(134),(234),$$
$$(567),(11,12,13)$$

等共 6 个方程。注意当 $i<j<k$ 时，由于

$$f_{ij}=f_{ik}=f_{jk}=0,$$

相应方程变为

$$x_{ij}x_{ik}+x_{ij}x_{jk}+x_{ik}x_{jk}+x_{ik}=0。$$

不难验证，由于(Ⅰ)$_f$ 的解答需采取前面所述形式，诸方程(Ⅱ)$_f$ 都自然满足，因之 K'_2 必能嵌入平面。

由此知方程组(Ⅰ)$_f$，(Ⅱ)$_f$ 的解答实质上只有以下两组(在无损于一般性的假定 $x_{12}=x_{13}=x_{14}=0$ 下)：

(1) $x_{12}=x_{13}=x_{14}=x_{23}=x_{24}=x_{34}=x_{8,10}=x_{12,13}=0,$

$x_{56}=x_{57}=x_{67}=x_{11,12}=x_{11,13}=x_{15,16}=1$。

(2) $x_{12}=x_{13}=x_{14}=x_{56}=x_{57}=x_{67}=x_{11,12}=x_{11,13}$

$=x_{15,16}=0,$

$x_{23}=x_{24}=x_{34}=x_{8,10}=x_{12,13}=1$。

因而 T' 在平面中只有两个不同的嵌入。从这两个 T' 的嵌入不难获得 K'_2 在平面中的所有不同嵌入方法。

不仅 K_1，K_2 不能嵌入平面，凡含有 K_1 型或 K_2 型子图的线图都不能嵌入平面，而且 Kuratowski 还证明了其逆，即下述定理(证明从略，可参阅 Kuratowski, Sur le problème *des* courbes

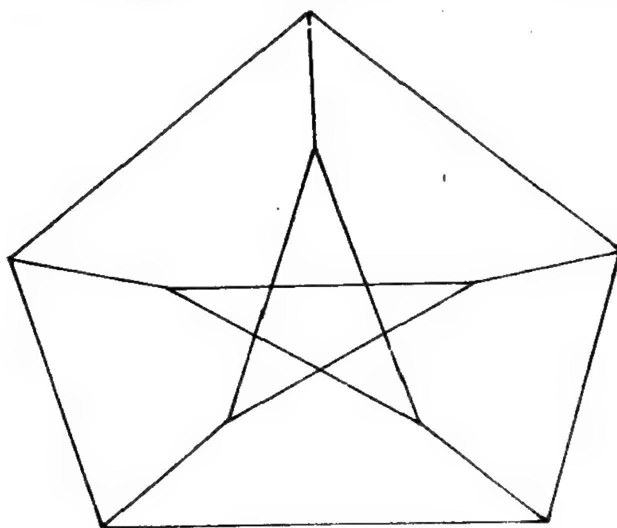


图 30

gauches en topologie, Fund. Math., 15(1930)271—283)。

Kuratowski 线图嵌入术 若一线图 G 含有一子图如 K_1 或 K_2，例如图 30 中以粗黑线表出的子图即为一 K_1 型线图，则上述 G 显然不能嵌入平面。反之，若 G 不含有任意 K_1 或 K_2 型的子图，则 G 必能嵌入平面，即

G 能嵌入平面 $\Leftrightarrow G$ 不含有 K_1 或 K_2 型子图

7. 线图嵌入第二基本定理

问题 Kuratowski 嵌入术给出了一个线图 G 能嵌入平面的充要条件，但这种判准并不是切实可行的，因为线图比较复杂时，难以判断是否有 K_1 或 K_2 型子图隐藏其中（参阅图 30)。另一方面，第 5 节 5 中的线图嵌入术（第一基本定理）给出了另一个切实可行的判准，把判断 G 能否嵌入平面归结为某方程组 $(\mathrm{I})_f$ 与 $(\mathrm{II})_f$ 是否有解答，并可由解答得出具体的具体嵌入方法。但方程组中 $(\mathrm{I})_f$ 是线性的，容易处理，$(\mathrm{II})_f$ 则是二次的，比较难办。试简化方程组以获得较简易的嵌入判准。

剖析 由 Kuratowski 嵌入术与实例的注释可知：设 G 中有一 K_1 或 K_2 型的子图 G'，在 G 中可选最大树形 T 以 G' 的最大树形 T' 为子树。不妨设 G，G' 都已改造使 T，T' 具有相同树根。任作 $f: T \subset$ 平面，并置 $f|T' \equiv f'$，则相应于 G，T 与 f 的方程组 $(\mathrm{I})_f$ 含有 G' 相应于 T' 与 f' 的方程组 $(\mathrm{I})_{f'}$ 为其子组。因 $(\mathrm{I})_{f'}$ 无解，故 $(\mathrm{I})_f$ 也无解。由此知，当 $(\mathrm{I})_f$ 有解时，G 必不能含有任一 K_1 或 K_2 型的子图。仍由 Kuratowski 嵌入术知，此时 G 必能嵌入平面。由第一基本定理知此时不仅 $(\mathrm{I})_f$ 有解，且 $(\mathrm{I})_f$，$(\mathrm{II})_f$ 联立起来的方程

组也必有解。但需注意，并不是（Ⅰ）$_f$ 的任一解答都可作为（Ⅰ）$_f$，（Ⅱ）$_f$ 联立起来的解。

总结上述，有下面的结论：

线图嵌入术（第二基本定理） 设线图 G，最大树形 T 与树根 O 如前。任作嵌入 $f: T \subset$ 平面，并作相应方程组

（Ⅰ）$_f$ $\qquad x_{\alpha\beta} = f_{\alpha\beta}$,

（Ⅱ）$_f$ $\qquad Q_{ijk}(f, x) = 1$,

则

G 能嵌入平面$\Leftrightarrow$（Ⅰ）$_f$ 有解

由此并知

（Ⅰ）$_f$ 有解$\Rightarrow$（Ⅰ）$_f$,（Ⅱ）$_f$ 有解

四、（平面性）线图的具体嵌入

1．问题说明与方法概述

问题 仍设线图 G 与最大树形 T 如前，其中 G 满足第一部分第 2 节中的限制条件，而 G，T 已改造如第三部分第 2 节中所述的，树根为 O。任作嵌入 $f: T \subset$ 平面。从 f 可作两方程组：

（Ⅰ）$_f$ $\qquad x_{\alpha\beta} = f_{\alpha\beta}$,

（Ⅱ）$_f$ $\qquad Q_{ijk}(f, x) = 1$。

由第三部分第 7 节中线图嵌入术（第二基本定理）已知方程组（Ⅰ）$_f$ 有解时，（Ⅰ）$_f$，（Ⅱ）$_f$ 联立起来的方程组也有解，因而依据第三部第 2 节中的线图嵌入术（第一基本定理），可改变 f 为另一嵌入 $g: T \subset$ 平面并扩充 g 为 G 的一个嵌入。但是，在建立

第三部分第 7 节中结果的过程中须借助于 Kuratowski 嵌入术，而后者只是一纯理论的结果，并非切实可行，为此提出下面的问题：重新建立第三部分第 7 节中的线图嵌入术，以便在解决上述问题时不依赖于 Kuratowski 的结果，且从$(\mathrm{I})_f$的一组解答获得$(\mathrm{I})_f$，$(\mathrm{II})_f$的解答时，应该依照确定步骤进行，因而切实可行，即可编成程序使用电子计算机以获得所需解答。

方法概述 为了解决上述问题，使得从一个只满足$(\mathrm{I})_f$的解出发能获得一个同时满足$(\mathrm{I})_f$，$(\mathrm{II})_f$的解，我们将把f以及$(\mathrm{I})_f$的解$\varepsilon=(\varepsilon_{ij})$逐步改变某些$\varepsilon_{ij}$的值，使满足$(\mathrm{II})_f$中的方程逐步增多，而改变后的$\varepsilon_{ij}$仍满足$(\mathrm{I})_f$。为此先引入以下的符号与概念。

设有f: $T\subset$平面与数组$\varepsilon=(\varepsilon_{ij})$。$e_a$是一外棱，两端为$a$，$a'$。在树形$T$中，连结$a$，$a'$与树根$O$的通道为$P_a$，$P_{a'}$，连结$a$，$a'$的通道为$P_a$。记$P_a$，$P_{a'}$的公共部分在$P_a$上的顶点为$v_a$，即$P_a\cap P_{a'}=P_{va}$。顶点$v_a$将称为$P_a$的**歧点**（有时也称为$e_a$的**歧点**），记$P_{va}$上以$v_a$为端点的树棱为$t_{va}$。

我们称$(f,\ \varepsilon)$对e_a**合格**，如果以下诸条件满足：

条件（1）对任一不同于e_a的外棱e_β，$\varepsilon=(\varepsilon_{ij})$满足$(\mathrm{I})_f$中与$(e_a,\ e_\beta)$相应的方程，即有$f_{a\beta}=\varepsilon_{a\beta}$。

条件（2）对以P_a上同一不等于v_a的顶点为前端的任三树棱t_i，t_j，t_k，若其中之一在P_a上，则$\varepsilon=(\varepsilon_{ij})$即满足$(\mathrm{II})_f$中与$(t_i,t_j,t_k)$相应的方程：$Q_{ijk}(f,x)=1$。

条件（3）对以P_a上同一顶点v_a为前端的任三树棱t_i，t_j，t_k，若其中两个在P_a上，则$\varepsilon=(\varepsilon_{ij})$即满足$(\mathrm{II})_f$中与$(t_i,t_j,t_k)$相应的方程：$Q_{ijk}(f,x)=1$。

如果数组$\varepsilon=(\varepsilon_{ij})$又是$(\mathrm{I})_f$的解答且满足上述三条件，则称$\varepsilon$是$f$下对$e_a$的**合格解**。

于是，我们的方法可概述如下。

将诸外棱排成一适当次序如下：

$$e_1,\ e_2,\ \cdots,\ e_a,\ \cdots,\ e_L,$$

从 f: $T\subset$平面以及（Ⅰ）$_f$ 的任一解 $\varepsilon=(\varepsilon_{ij})$ 出发，逐步改变为数组

$$\varepsilon^1=(\varepsilon^1_{ij}),\varepsilon^2=(\varepsilon^2_{ij}),\cdots,\varepsilon^a=(\varepsilon^a_{ij}),\cdots,\varepsilon^L=(\varepsilon^L_{ij}),$$

使每一 ε^a 都是(Ⅰ)$_f$ 的解，而(f,ε^1)对 e_1 合格，(f,ε^2)除对 e_1 合格外，又对 e_2 合格，依此类推。最后，所得 $\varepsilon^L=(\varepsilon^L_{ij})$即为同时满足(Ⅰ)$_f$,(Ⅱ)$_f$ 的一组解答。

方法的改进 上述方法中嵌入 f: $T\subset$平面是固定的，只是逐步改变数组 $\varepsilon=(\varepsilon_{ij})$ 的值。但在 f 下诸棱的相互位置与 ε 及其改变之间并无显著关系。为此有必要不仅改变 ε，且改变嵌入 f，使改变后的嵌入 g 与数组 $\eta=(\eta_{ij})$ 满足 $(f,\ \varepsilon)$ 所满足的那些相应的方程，而 g 与 η 间却有比较简单直观的相互关系。这可以帮助解决原来的问题。

为此先引入以下的符号与概念。

设将 f: $T\subset$平面改变为 g: $T\subset$平面。对于（Ⅰ）$_f$ 的数组 $\varepsilon=(\varepsilon_{ij})$，有（Ⅰ）$_g$ 的一个**相应**数组 $\eta=(\eta_{ij})$，使

$$\eta_{ij}+g_{ij}=\varepsilon_{ij}+f_{ij}。$$

此时我们简记作

$$(f,\ \varepsilon)\sim(g,\ \eta)。$$

注意 $\varepsilon_{ij}=\varepsilon_{ji}$, $\eta_{ij}=\eta_{ji}$, 而 $f_{ij}={}_{ji}+1$, $g_{ij}=g_{ji}+1$。由于对任意外棱 e_α, e_β, 依据第三部分第 5 节引入的 $\varepsilon_{\alpha\beta}$, $\eta_{\alpha\beta}$, 显然满足

$$\eta_{\alpha\beta}+g_{\alpha\beta}=\varepsilon_{\alpha\beta}+f_{\alpha\beta},$$

故有以下断言：

设$(f,\varepsilon)\sim(g,\eta)$。

(1) 若 $\varepsilon=(\varepsilon_{ij})$是(Ⅰ)$_f$ 的解，则 $\eta=(\eta_{ij})$也是(Ⅰ)$_g$ 的解。

(2) 对以同一顶点为前端的任三树棱 t_i, t_j, t_k, 若 ε 满

足$(\mathrm{II})_f$中与(t_i, t_j, t_k)相应的方程，则 η 也满足$(\mathrm{II})_g$中与(t_i, t_j, t_k)相应的方程。

由此不难看出：

(3) 若(f, ε)对某一外棱 e_a 合格，则(g, η)也对 e_a 合格。

(4) 若 ε 是f 下对e_a 的合格解，则 η 也是g 下对e_a 的合格解。

由于 (1) — (4)，可将原来方法改进为以下更一般的形式：

仍将诸外棱排成一适当次序

$$e_1, \quad e_2, \quad \cdots, \quad e_\alpha, \quad \cdots, \quad e_L,$$

从 $f: T\subset$平面以及$(\mathrm{I})_f$的任一解$\varepsilon=(\varepsilon_{ij})$出发，逐步改变 f 为嵌入

$$g^1, \quad g^2, \quad \cdots, \quad g^\alpha, \quad \cdots, \quad g^L: T\subset \text{平面},$$

以及 ε 为数组

$$\eta^1=(\eta^1_{ij}), \eta^2=(\eta^2_{ij}), \cdots, \eta^\alpha=(\eta^\alpha_{ij}), \cdots, \eta^L=(\eta^L_{ij}),$$

使每一 η^α 都是$(\mathrm{I})_{g\alpha}$的解，而(g^1, η^1)对 e_1 合格，(g^2, η^2)除对 e_1 合格外，又对 e_2 合格，依次类推。最后所得 $\eta^L=(\eta^L_{ij})$即为同时满足$(\mathrm{I})^L_g$，$(\mathrm{II})^L_g$ 的一组解答。作数组 $\varepsilon^L=(\varepsilon^L_{ij})$，使

$$(f, \varepsilon^L)\sim(g, \eta^L),$$

则 ε^L 也是同时满足$(\mathrm{I})_f$，$(\mathrm{II})_f$的一组解答。

以下各节将解释逐次改变嵌入与数组的细节，也将说明诸外棱排成适当次序的方法。

2. 旋数的改变

问题 上节提出的方法主要在于改变一个数组 $\varepsilon=(\varepsilon_{ij})$ 的值使满足 $(\mathrm{II})_f$ 的方程个数越来越多，为此，试检查改变前后的数组对满足 $(\mathrm{II})_f$ 与否的影响。

剖析 对第二部分第 4 节中的树形比嵌术中从嵌入 $f: T \subset$平面的旋数组（f_{ij}）改变为另一嵌入的旋数组时，引入了改变值的数组（ε_{ij}）。由于对以同一顶点 v 为前端的任三树棱 t_i，t_j，t_k，数组（ε_{ij}，ε_{ik}，ε_{jk}）可取的值有 8 种，而将 t_i，t_j，t_k 在 f 下的排列次序改变时的方式却只有 6 种，这说明作为旋数改变的（ε_{ij}，ε_{ik}，ε_{jk}）三数是不能任意的。其间的制约关系由（Ⅱ）$_f$ 中的方程

$$Q_{ijk}(f,x)=1$$

表达，即三数须满足这一方程。由这一方程的几何意义容易解答上述问题。

为此，不妨设在 f 下从 t_v 起依反钟向排列的次序为 t_v，t_i，t_j，t_k，于是有下表：

改变后反钟向排列次序	ε_{ij}	ε_{ik}	ε_{jk}
t_v, t_i, t_j, t_k	0	0	0
t_v, t_i, t_k, t_j	0	0	1
t_v, t_j, t_k, t_i	1	1	0
t_v, t_j, t_i, t_k	1	0	0
t_v, t_k, t_i, t_j	0	1	1
t_v, t_k, t_j, t_i	1	1	1

由表知，$(\varepsilon_{ij},\varepsilon_{ik},\varepsilon_{jk})$不能取的值为$(0,1,0)$与$(1,0,1)$，且知从这两组改变三数中的任意一个或两个，所得新数组即与表中某一确定排列的数组相当。这一结论显然与原先 f 下 t_v，t_i，t_j，t_k 的排列次序无关，因而有下述结论：

旋数改变术 设有嵌入 $f: T\subset$平面，数组 $\varepsilon=(\varepsilon_{ij})$，以及以同一顶点为前端的三树棱 t_i，t_j，t_k，若三数 ε_{ij}，ε_{ik}，ε_{jk} 不满足（Ⅱ）$_f$ 中与（t_i，t_j，t_k）相应的方程：Q_{ijk}（f，x）

$=1$，则任意改变其中一个或两个的值，新得到的三数 ϵ'_{ij}，ϵ'_{ik}，ϵ'_{jk}即能满足这一方程。

3. 树形嵌入的调整

问题　依据第一节中的要求，试将嵌入 $f: T \subset$ 平面改变为嵌入 $g: T \subset$ 平面，又相应改变数组 $\epsilon=(\epsilon_{ij})$ 为数组 $\eta=(\eta_{ij})$，使 $(f,\epsilon)\sim(g,\eta)$ 而在 g 下 η 较为简单易于处理。

改变方法　改变将对一确定的外棱 e_α 进行。设 P_α 上以歧点 v_α 为前端的两树棱为 t_p，t_q。必要时交换 p，q 记号，故恒可假设 $f_{pq}+\epsilon_{pq}=0$。又简记 $t_{v\alpha}$ 为 t_0.

若 $\epsilon_{pq}=1$，则改变 f 为 g 时，将交换 t_p，t_q 的位置，否则 t_p，t_q 的相互位置保持不变。以下设这一步骤已首先完成，故不妨设在 g 下与在 f 下 t_0，t_p，t_q 都依反钟向排列，于是有

$$g_{pq}=0,\quad \eta_{pq}=0。$$

设 v' 为 P_α 上或 $P_{v\alpha}$ 上任一不同于歧点 v_α 的顶点，t_r 为 P_α 或 $P_{v\alpha}$ 上以 v' 为前端的树棱，而 t_k 为以 v' 为前端但不在 P_α 或 $P_{v\alpha}$ 上的任一树棱。若 $\epsilon_{kr}=0$，则在 g 下将保持 f 下 t_k，t_r 的旋向。否则改变旋向，即从 f 改变为 g 时，t_k 将移置 t_r 的另一侧。于是不论何时，都有

$$\eta_{kr}=0。$$

次设 t_k 为以歧点 v_α 为前端而不同于 t_p，t_q 的树棱。若 t_k 在 f 下位于 t_0 与 t_p 之间，即在 f 下依反钟向排列次序为 t_0 $(=t_{v\alpha})$，t_k，t_p，t_q 时，将改变 f 为 g 使在 g 下 t_k 的位置依次如下：

(1) $(\epsilon_{kp},\epsilon_{kq})=(0,0)$ 时，t_k 在 g 下仍在 t_0，t_p 之间。

(2) $(\epsilon_{kp},\epsilon_{kq})=(1,1)$ 时，t_k 在 g 下位于 t_0，t_q 之间。

(3)，(4) $(\epsilon_{kp},\ \epsilon_{kq})=(1,\ 0)$ 或 $(0,\ 1)$ 时，t_k 在 g

下位于 t_p， t_q 之间。

参阅附图 31，其中波状线为 t_k 改变后在 g 下的位置。

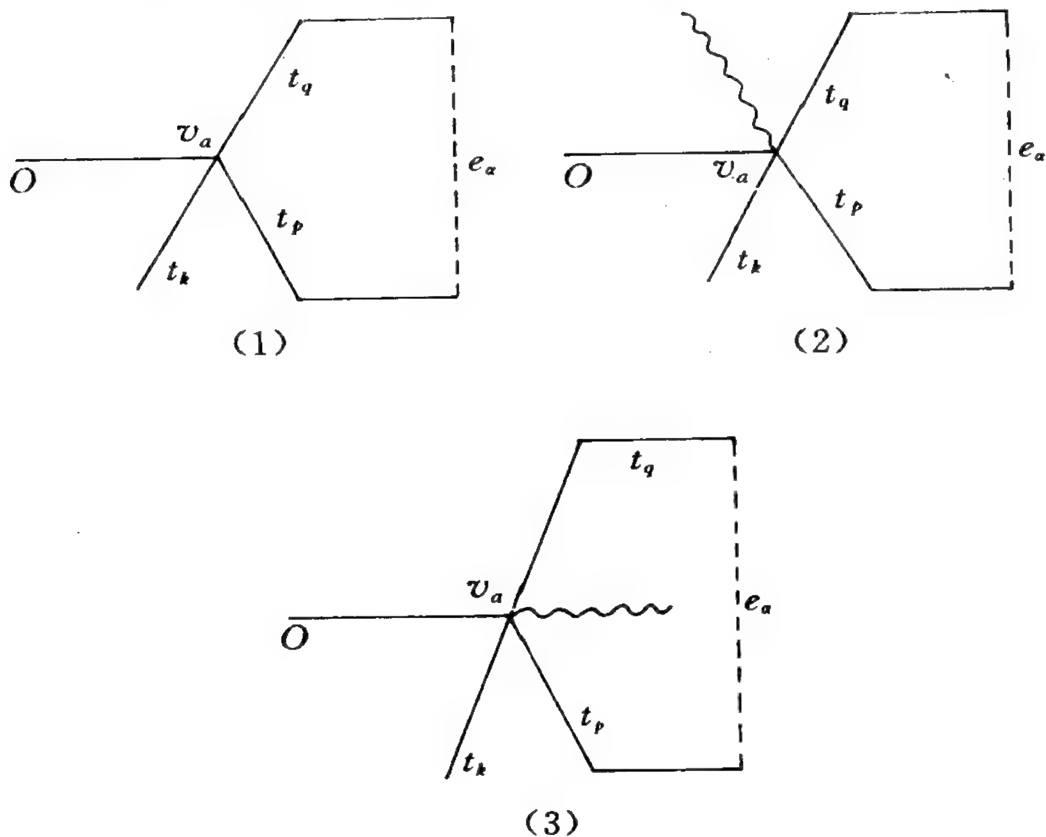


图 31

若 t_k 仍为以歧点 v_a 为前端而不同于 t_p, t_q 的树棱,但在 f 下位于 t_0 与 t_q 之间，则改变方法与上相似。若在 f 下位于 t_p, t_q 之间,则在$(\epsilon_{kp},\epsilon_{kq})=(0,0)$或$(1,1)$时,在 g 下 t_k 仍置于 t_p, t_q 之间,而在$(\epsilon_{kp},\epsilon_{kq})=(0,1)$或$(1,0)$时, t_k 将依次移置 t_0, t_q 与 t_0, t_p 之间。

最后，设顶点 v'既不在 P_a 上也不在 P_{va}上，则以 v'为前端的诸棱间的旋向将保持不变，即对任两以 v'为前端的树棱 t_i， t_j，有

$$g_{ij}=f_{ij}, \quad \eta_{ij}=\epsilon_{ij}。$$

由于对以 P_a 或P_{va}上顶点为前端而不在P_a 与P_{va}上的任两

树棱，其相互位置改变与否除已可归结为以上改变者外并无明确规定，故依据以上诸规则所得嵌入 g 不必是唯一的，但任一这样改变所得的（g，η）都将称为（f，ε）对 e_α 的一个调整，简记为

$$(f,\varepsilon)\xrightarrow[e_\alpha]{\sim}(g,\eta)。$$

易见经嵌入调整所得 g,η 具有以下简单特征。

嵌入调整术　设$(f,\varepsilon)\xrightarrow[e_\alpha]{\sim}(g,\eta)$，则

（1）若 v'是P_α 或$P_{v\alpha}$上$\neq$歧点 v_α 的顶点，t_r 是 P_α 或 $P_{v\alpha}$ 上以 v'为前端的树棱，而 t_k 为以 v'为前端的另一树棱，则

$$\eta_{kr}=0。$$

（2）若 t_p，t_q 是 P_α 上以 v_α 为前端的两树棱，而 $t_{v\alpha}$，t_p，t_q 在 g 下依反钟向排列，则

$$\eta_{pq}=0。$$

（3）若 t_p，t_q 如前，而 t_k 为以 v_α 为前端而不同于 t_p，t_q 的任一树棱，则

在 g 下 t_k 位于 t_0，t_p 或 t_0，t_q 之间时，都有

$$\eta_{kp}=\eta_{kq}=0。$$

在 g 下 t_k 位于 t_p，t_q 之间时，或

$$\eta_{kp}=\eta_{kq}=0$$

或

$$\eta_{kp}=\eta_{kq}=1。$$

（4）在(3)中除最后情形外，η_{kp}，η_{kq}，η_{pq}三数都满足$(\mathrm{II})_g$中与(t_k,t_p,t_q)相应的方程，亦即原来数组 $\varepsilon=(\varepsilon_{ij})$满足$(\mathrm{II})_f$中与$(t_k,t_p,t_q)$相应的方程。

（5）若$e_\alpha=\widehat{aa'}$，b 是 T 的任一不同于α，α'的自由端，则由

(1)—(4)得，$\eta_{ab}=\eta_{a'b}=0$ 或 $\eta_{ab}=\eta_{a'b}=1$，而后者只在 P_b 通过 v_a 且在 g 下 t_p, t_q 为 P_b 所隔开的情形下始能出现。因而不论何时，都有

$$\eta_{ab}=0。$$

(6) 由 (5) 知，对任一不同于 e_α 的外棱 e_β，都有

$$\eta_{\alpha\beta}=0。$$

4. 方程组 $(\mathrm{I})_f$ 解答的调整

问题 设有 $f: T\subset$平面，$(I)_f$ 的解 $\varepsilon=(\varepsilon_{ij})$，以及外棱 e_α，求将 ε 改变为数组 $\varepsilon'=(\varepsilon'_{ij})$，使 (f, ε') 对 e_α 合格，且 ε' 仍满足方程组 $(I)_f$，因而 ε' 为 f 下对 e_α 的合格解。

数组调整术 依以下规则可将数组 $\varepsilon=(\varepsilon_{ij})$ 改变为数组 $\varepsilon'=(\varepsilon'_{ij})$，使 ε 是 $(I)_f$ 的解答时，ε' 为 f 下对 e_α 的合格解。

规则(1)设 v' 为 P_α 上任一不同于歧点 v_α 的顶点，t_r 为 P_α 上以 v' 为前端的树棱，t_i，t_j 为以 v' 为前端而不同于 t_r 的任两树棱，若 ε 不满足 $(II)_f$ 中与 (t_r, t_i, t_j) 相应的方程，则改变 ε_{ij} 的值为

$$\varepsilon'_{ij}=\varepsilon_{ij}+1。$$

否则置

$$\varepsilon'_{ij}=\varepsilon_{ij}。$$

此外，又置

$$\varepsilon'_{ir}=\varepsilon_{ir}，\ \varepsilon'_{jr}=\varepsilon_{jr}。$$

规则(2)设 t_p，t_q 为 P_α 上以歧点 v_α 为前端的两树棱，t_k 为以 v_α 为前端而不同于 t_p，t_q 的任一树棱。若 ε 不满足 $(II)_f$ 中与 (t_k, t_p, t_q) 相应的方程，则改变 ε_{kp}，ε_{kq} 为

$$\varepsilon'_{kp}=\varepsilon_{kp}+1，\ \varepsilon'_{kq}=\varepsilon_{kq}+1。$$

否则仍置

$$\varepsilon'_{kp}=\varepsilon_{kp},\quad \varepsilon'_{kq}=\varepsilon_{kq}。$$

此外，又置

$$\varepsilon'_{pq}=\varepsilon_{pq}。$$

规则(3)设 t_p，t_q 如规则 2，而 t_i，t_j 为以 v_α 为前端而不同于 t_p，t_q 的任两树棱。若 $\varepsilon=(\varepsilon_{ij})$ 恰能满足 $(\mathrm{II})_f$ 中与 (t_i, t_j, t_p)，(t_i, t_j, t_q) 相应的两方程之一，则改变 ε_{ij} 为

$$\varepsilon'_{ij}=\varepsilon_{ij}+1。$$

否则置

$$\varepsilon'_{ij}=\varepsilon_{ij}。$$

规则(4)设 t_i，t_j 为以不在 P_α 上的同一顶点为前端的任两树棱，则置

$$\varepsilon'_{ij}=\varepsilon_{ij}。$$

以上由数组 $\varepsilon=(\varepsilon_{ij})$ 依据规则 1—4 改变为数组 $\varepsilon'=(\varepsilon'_{ij})$ 的方法称为在 f 下对 e_α 的**数组调整**，简记为

$$(f, \varepsilon)\xrightarrow[e_\alpha]{}(f, \varepsilon')。$$

在 $\varepsilon=(\varepsilon_{ij})$ 是 $(\mathrm{I})_f$ 的解答时，如此所得数组 $\varepsilon'=(\varepsilon'_{ij})$ 即为 f 下对 e_α 的一个合格解。

注析. 为验证术文中的结论，先改变 (f, ε) 为较易处理的嵌入 g 与数组 η，即依嵌入调整术作

$$(f, \varepsilon)\xrightarrow[e_\alpha]{\sim}(g, \eta)。$$

命 $\eta'=(\eta'_{ij})$ 为 f 改变为 g 时与 ε' 相应的数组，即

$$(f, \varepsilon')\sim(g, \eta'),$$

于是 η' 是 η 在 g 下对 e_α 的数组调整，即

$$(g, \eta)\xrightarrow[e_\alpha]{}(g, \eta')。$$

而与术文中由 ε 改变为 ε' 的诸规则相应，η 改变为 η' 所遵循的规则可重述如下：

规则(1′)设 v' 为 P_a 上任一不同于 v_a 的顶点，t_r 为 P_a 上以 v' 为前端的树棱，t_i，t_j 为以 v' 为前端的任两不同于 t_r 的树棱。若 η 不满足 $(\mathrm{II})_g$ 中与（t_r，t_i，t_j）相应的方程，即 t_i，t_j 在 g 下位于 t_r 的两侧，且 $\eta_{ij}=1$，则改变 η_{ij} 的值为

$$\eta'_{ij}=\eta_{ij}+1。$$

否则置

$$\eta'_{ij}=\eta_{ij}。$$

此外，又置

$$\eta'_{ir}=\eta_{ir}，\quad \eta'_{jr}=\eta_{jr}，$$

即

$$\eta'_{ir}=0，\quad \eta'_{jr}=0。$$

规则(2′)设 t_p，t_q 为 P_a 上以歧点 v_a 为前端的两树棱，而在 g 下 t_0（$=t_{va}$），t_p，t_q 依反钟向排列，t_k 为以 v_a 为前端而不同于 t_p，t_q 的任一树棱。若 η 不满足 $(\mathrm{II})_g$ 中与（t_k，t_p，t_q）相应的方程，即 t_k 在 g 下位于 P_a 的与 O 不同的一侧，且 $\eta_{kp}=\eta_{kq}=1$，则改变 η_{kp}，η_{kq} 为

$$\eta'_{kp}=\eta_{kp}+1，\quad \eta'_{kq}=\eta_{kq}+1。$$

否则仍置

$$\eta'_{kp}=\eta_{kp}，\quad \eta'_{kq}=\eta_{kq}。$$

此外，又置

$$\eta'_{pq}=\eta_{pq}，$$

即

$$\eta'_{pq}=0。$$

规则(3′)设 t_p，t_q 如规则（2′），而 t_i，t_j 为以 v_a 为前端但不同于 t_p，t_q 的任两树棱，若 $\eta=(\eta_{ij})$ 恰能满足 $(\mathrm{II})_g$ 中（t_i，t_j，t_p），（t_i，t_j，t_q）相应的两方程之一，即有下两情形之一，则改变 η_{ij} 为

$$\eta'_{ij} = \eta_{ij} + 1。$$

否则置

$$\eta'_{ij} = \eta_{ij}。$$

两情形如下：

①t_i，t_j 在 g 下位于 t_p，t_q 之间，或

$$\eta_{ip} = \eta_{iq} = 1, \quad \eta_{jp} = \eta_{jq} = 0,$$

或

$$\eta_{ip} = \eta_{iq} = 0, \quad \eta_{jp} = \eta_{jq} = 1。$$

②t_i，t_j 在 g 下与 t_p，t_q 相隔离，且有

$$\eta_{ij} = 1。$$

规则(4′)设 t_i，t_j 为以不在 P_a 上的同一顶点为前端的任两树棱，则置

$$\eta'_{ij} = \eta_{ij}。$$

根据 g，η、η'的作法不难验证以下诸简单性质：

性质（1）以 P_a 或 P_{va} 上同一顶点为前端的任两树棱 t_i，t_j，只需至少有一在 P_a 或 P_{va} 上，即有

$$\eta'_{ij} = 0。$$

性质(2)设 T 的两自由端 c，d 在 g 下位于 P_a 的异侧，则必有

$$\eta'_{cd} = 0。$$

性质(3)设 T 的两自由端 c，d 与树根 O 在 g 下都位于 P_a 的同侧，则必有

$$\eta'_{cd} = \eta_{cd}。$$

性质(4)设自 T 的两自由端 c，d 至树根 O 的通道 P_c，P_d 首次相遇的顶点不在 P_a 上，则必有

$$\eta'_{cd} = \eta_{cd}。$$

性质(5)设树棱 t_i，t_j，t_k 均以 v_a 为前端，且三者或为

t_p，t_q，或在 g 下位于 P_α 的与 O 不同的一侧，则在（η'_{ij}，η'_{ik}，η'_{jk}）三数中与相应三数（η_{ij}，η_{ik}，η_{jk}）不相等的个数或为 0，或为 2。

今设 ε 是 $(\mathrm{I})_f$ 的解，验证（f，ε'）对 e_α 合格如下。

首先，依据第 2 节旋数改变术以及改变 ε 为 ε' 的规则 (1) — (3)，可见（f，ε'）满足对 e_α 合格的条件（2）与 (3)。

为验证（f，ε'）也满足对 e_α 合格的条件（1）。可注意依第 3 节中的嵌入调整术，对任一不同于 e_α 的外棱 e_β 有 $\eta_{\alpha\beta}=0$。由于 ε 是 $(\mathrm{I})_f$ 的解，而（f，ε）～（g，η），故 η 也是 $(\mathrm{I})_g$ 的解，特别需满足 $(\mathrm{I})_g$ 中与（e_α，e_β）相应的方程，故有 $g_{\alpha\beta}=\eta_{\alpha\beta}=0$。由关于（$g$，$\eta$，$\eta'$）的性质 1，显然有（$e_\beta=\widehat{bb'}$）：

$$\eta'_{ab}=\eta'_{a'b}=\eta'_{ab'}=\eta'_{a'b'}=0,$$

故 $\eta'_{\alpha\beta}=0$。因而 $g_{\alpha\beta}=0$ 也可写成 $g_{\alpha\beta}=\eta'_{\alpha\beta}$，即 η' 满足 $(\mathrm{I})_g$ 中与（e_α，e_β）相应的方程。于是 ε' 也满足 $(\mathrm{I})_f$ 中与（e_α，e_β）相应的方程，即（f，ε'）满足对 e_α 合格的条件 1。

由以上验证并知：若 ε 是 $(\mathrm{I})_f$ 的解或 η 是 $(\mathrm{I})_g$ 的解，则（g，η，η'）还具有以下性质。

性质(6)对任一不同于 e_α 的外棱 e_β，有

$$g_{\alpha\beta}=\eta'_{\alpha\beta}=0,$$

因而 e_β 的两端在 g 下必位于 P_α 的同侧。

我们称从（f，ε）依据嵌入调整术获得的嵌入 g 以及任一数组 $\eta'=$（η'_{ij}）对 e_α **简单合格**，如果 η' 对 e_α 合格且满足前面性质（1），（2）与（6），特别是在 ε 满足方程组 $(\mathrm{I})_f$ 时，嵌入 g 以及由诸规则（1′）—（4′）所得数组 η' 对 e_α 简单合格。

术文中 ε 满足 $(\mathrm{I})_f$ 时（f，ε）对 e_α 合格这一部分已验

证完毕。其次，需再验证在 ε 满足 $(\mathrm{I})_f$ 时 ε' 也满足 $(\mathrm{I})_f$。或只需再验证在 η 满足 $(\mathrm{I})_g$ 时，η' 也满足 $(\mathrm{I})_g$，即对任两外棱 e_β，e_γ，η' 必满足 $(\mathrm{I})_g$ 中与 (e_β, e_γ) 相应的方程。

在 e_β，e_γ 中有一为 e_α 时，已知这一结论成立，故可设 e_β，e_γ 都不同于 e_α。以下就各种不同情形分别讨论之。

(1) 设 e_β，e_γ 中有一，例如 e_β，其歧点 v_β 不在 P_α 上也不在 $P_{v\alpha}$ 上。命 e_β 的两端为 b 与 b'：$e_\beta=\widehat{bb'}$，而 c 为 T 的任一不同于 b，b' 的自由端，则当 P_c 不经过 v_β 时，显然有

$$\eta_{bc}=\eta_{b'c},\quad \eta'_{bc}=\eta'_{b'c}。$$

当 P_c 经过 v_β 时，b，b'，c 必都在 P_α 同侧，故由关于 (g, η, η') 的性质 3，有

$$\eta'_{bc}=\eta_{bc},\quad \eta'_{b'c}=\eta_{b'c}。$$

不论何时都有 $\eta'_{\beta c}=\eta_{\beta c}$，因而

$$\eta'_{\beta\gamma}=\eta_{\beta\gamma}。$$

由于 η 满足 $(\mathrm{I})_g$ 中与 (e_β, e_γ) 相应的方程，即有 $g_{\beta\gamma}=\eta_{\beta\gamma}$，故 $g_{\beta\gamma}=\eta'_{\beta\gamma}$，即 η' 也满足 $(\mathrm{I})_g$ 中与 (e_β, e_γ) 相应的方程。

(2) 设 e_β，e_γ 中有一，例如 e_β，其两端 b，b' 在 g 下与树根 O 都位于 P_α 的同侧，则当 e_γ 的两端 c，c' 在 g 下与 O 位于 P_α 的同侧时，由 (g, η, η') 的性质 3，有

$$\eta'_{bc}=\eta_{bc},\quad \eta'_{bc'}=\eta_{bc'},\quad \eta'_{b'c}=\eta_{b'c},\quad \eta'_{b'c'}=\eta_{b'c'},$$

故 $\eta'_{\beta\gamma}=\eta_{\beta\gamma}$。与 (1) 相同，知 η' 满足 $(\mathrm{I})_g$ 中与 (e_β, e_γ) 相应的方程。反之，当 c，c' 在 g 下与 O 位于 P_α 的异侧时，由关于 (g, η, η') 的性质 (2)，有

$$\eta'_{bc}=\eta'_{bc'}=\eta'_{b'c}=\eta'_{b'c'}=0。$$

因而 $\eta'_{\beta\gamma}=0$。另一方面，又显然有 $g_{\beta\gamma}=0$，故 $g_{\beta\gamma}=\eta'_{\beta\gamma}$ $(=0)$，即 η' 仍满足 $(\mathrm{I})_g$ 中与 (e_β, e_γ) 相应的方程。

(3) 设 e_β，e_γ 诸端点在 g 下都位于 P_α 的与 O 不同的一

侧，且 P_β，P_γ 的歧点都在 P_α 上，又设歧点之一，例如 $v_\beta \neq v_\alpha$。命 $e_\beta = \widehat{bb'}$，$e_\gamma = \widehat{cc'}$，易见或

$$\eta'_{bc} = \eta_{bc}, \quad \eta'_{b'c} = \eta_{b'c},$$

或

$$\eta'_{bc} = \eta'_{b'c}, \quad \eta_{bc} = \eta_{b'c},$$

因而不论何时都有 $\eta'_{\beta c} = \eta_{\beta c}$。同样 $\eta'_{\beta c'} = \eta_{\beta c'}$，故 $\eta'_{\beta\gamma} = \eta_{\beta\gamma}$。与（1）相同，知 η' 仍满足 $(\mathrm{I})_g$ 中与（e_β，e_γ）相应的方程。

（4）最后设 $e_\beta = \widehat{bb'}$，$e_\gamma = \widehat{cc'}$ 的诸端点在 g 下仍位于 P_α 的与 O 不同的一侧，但 P_β，P_γ 的歧点都在 v_α 处：$v_\beta = v_\gamma = v_\alpha$。若自 b，c 至 O 的通道首次相遇的顶点不在 v_α 处，则由关于（g，η，η'）的性质（1），（4），有

$$\eta'_{bc} = \eta_{bc}, \quad \eta'_{b'c} = \eta'_{bb'}, \quad \eta_{b'c} = \eta_{bb'}。$$

因而有

$$\eta'_{bc} + \eta'_{b'c} + \eta'_{bb'} = \eta_{bc} + \eta_{b'c} + \eta_{bb'}。$$

自 b'，c 至 O 的通道首次相遇的顶点不在 v_α 时也同样。若自 b，c 至 O 以及自 b'，c 至 O 的通道首次相遇的顶点都是 v_α，则由性质 5，仍得上式。同样不论何时都有

$$\eta'_{bc'} + \eta'_{b'c'} + \eta'_{bb'} = \eta_{bc'} + \eta_{b'c'} + \eta_{bb'}。$$

二式相加，即得

$$\eta'_{\beta\gamma} = \eta_{\beta\gamma}。$$

于是与（1）同样知 η' 仍满足 $(\mathrm{I})_g$ 中与（e_β，e_γ）相应的方程。

至此，已知在各种情形下 $\eta' = (\eta'_{ij})$ 都满足 $(\mathrm{I})_g$ 中所有方程，调整术验证完毕。

5．线图嵌入第三基本定理

问题 上节数组调整术给出了从任一 $(\mathrm{I})_f$ 的解答 $\varepsilon =$

(ε_{ij}) 改变成对一固定外棱 e_α 合格解的方法。在应用第一节中的方法从 ε 获得同时满足 $(\mathrm{I})_f$，$(\mathrm{II})_f$ 的解答时，关键的一点在于诸外棱需排列成一适当次序，使依调整术逐次改变数组时，对先前诸棱的合格性不致因新的改变而丧失。试求这种能符合要求的适当排列次序。

剖析 为求得适当的外棱排列次序，需先就对两个外棱处于较简单关系的情况下，调整对合格性的影响并进行较细致的分析。为此考虑外棱 e_α，e_β，与 P_α，P_β 的歧点 v_α，v_β。假设下两条件能满足：

(1) v_α 在通道 $P_{v\beta}$ 上。

(2) P_β 与 P_α 至少有一树棱公共。

今设已知嵌入 $g: T\subset$ 平面与 $(\mathrm{I})_g$ 的解 $\eta=(\eta_{ij})$ 对 e_α 简单合格，试考虑依调整术

$$(g,\ \eta)\ \xrightarrow[e_\beta]{\sim}\ (h,\ \zeta)\ \xrightarrow[e_\beta]{}\ (h,\ \zeta')$$

以获得对 e_β 简单合格的嵌入 $h: T\subset$ 平面与数组 $\zeta'=(\zeta'_{ij})$ 时对 e_α 合格性的影响。

由于 P_α，P_β 至少有一树棱公共，且 $(g,\ \eta)$ 对 e_α 简单合格，可知依据嵌入调整术以获得 h 时，交换 g 下 P_β 上以 v_β 为前端的两树棱位置这一步是不必要的。

以下将对嵌入与数组的各改变步骤逐一检查，注意只有下面两种改变可能引起丧失对 e_α 的合格性：

(1) 顶点 w 在 P_α 或 $P_{v\alpha}$ 上，树棱 t_k 以 w 为前端而不在 P_α 或 $P_{v\alpha}$ 上，由 g 调整为 h 时，t_k 在 g 下对 P_α 或 $P_{v\alpha}$ 所位侧向的改变。

(2) 顶点 w 在 P_α 上，树棱 t_i，t_j 以 w 为前端，由 ζ 调整为 ζ' 时，ζ_{ij} 的改变。

下面检查这两种改变的影响。

(1)′因（g，η）已对 e_α 简单合格，故对前端在 P_α 或 $P_{v\alpha}$ 上的任一树棱 t_k，t_k 在 g 下位于 P_α 的某一侧时，在 h 下仍将保持在这一侧。

(2)′从上节数据调整术的注析中，关于（g，η，η'）的性质 6 已知 e_β 的两端在 g 下，因而也在 h 下，必须位于 P_α 的同侧。又知（g，η）对 e_α 简单合格，故对前端 w 同在 P_α 上的任两树棱 t_i，t_j，在必须改变 ζ_{ij} 为 $\zeta_{ij}=\zeta_{ij}+1$ 时，应有

$w=P_\alpha$，P_β 公共部分的两端之一；

t_i，t_j 与 e_β 两端在 h 下位于 P_α 同侧；

t_i，t_j 在 h 下位于 P_β 的异侧；

$\zeta_{ij}=1$。

由以上分析易见，所得（h，ζ'）不仅对 e_β 简单合格，且对 e_α 也依然保持简单合格。

同样可知，设 e_α，e_β 如前，而对另一外棱 e_γ，P_r 的歧点不在 P_β 上而原来（g，η）对 e_r 简单合格，则（h，ζ'）也依然对 e_γ 简单合格。

由此易得排列诸外棱的一种合适方法，以为第一节中所说改进方法之用。

由第一部分第 2 节中关于线图 G 的限制，可将诸外棱 e_α 排成次序

$$e_1,\ e_2,\ \cdots,\ e_\alpha,\ \cdots,\ e_L,$$

使对每一外棱 e_α，$\alpha>1$，有下二性质：

(1) P_α 至少与 $\beta<\alpha$ 的某一 P_β 有一树棱公共。

(2) 若外棱 e_β 的歧点 v_β 在 $P_{v\alpha}$ 上且 $\neq v_\alpha$，则 e_β 必已出现于 e_1，e_2，…，$e_{\alpha-1}$ 中。

今从嵌入 $f: T\subset$ 平面与满足 $(\mathrm{I})_f$ 的解答 $\varepsilon=(\varepsilon_{ij})$ 出发，依据嵌入调整术与数组调整术逐步改变（f，ε）为

$$(f,\ \varepsilon)\xrightarrow[e_1]{}(g^1,\ \eta^1)\sim(f,\ \varepsilon^1),$$

$$(g^1,\ \eta^1)\xrightarrow[e_2]{}(g^2,\ \eta^2)\sim(f,\ \varepsilon^2),$$

$$\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots$$

$$(g^{L-1},\ \eta^{L-1})\xrightarrow[e_L]{}(g^L,\ \eta^L)\sim(f,\ \varepsilon^L)。$$

于是，最后所得 ε^L 同时满足 $(\mathrm{I})_f$，$(\mathrm{II})_f$ 两方程组。

这给出了线图嵌入第二基本定理的又一证明，而这一证明不仅不再依赖于 Kuratowski 嵌入术，而且由于每一 ε^α 都系在 f 下从 $\varepsilon^{\alpha-1}$（$\varepsilon^0=\varepsilon$）依数据调整术中规则（1）—（4）获得，因而切实可行，不难使之机器化。因之可将以前的嵌入术加强成下面形式。

线图具体嵌入术（第三基本定理） 设有线图 G，最大树形 T 如前，任作嵌入 $f: T\subset$平面，设 $\varepsilon=(\varepsilon_{ij})$ 是方程组

$(\mathrm{I})_f$ $$x_{\alpha\beta}=f_{\alpha\beta}$$

的任一解答，则可将诸外棱 e_α 排列成一适当次序

$$e_1,\ e_2,\ \cdots,\ e_\alpha,\ \cdots,\ e_L,$$

并可作数组

$$\varepsilon^1=(\varepsilon_{ij}^1),\varepsilon^2=(\varepsilon_{ij}^2),\cdots,\varepsilon^\alpha=(\varepsilon_{ij}^\alpha),\cdots,\varepsilon^L=(\varepsilon_{ij}^L),$$

使每一 ε^α 在 f 下对 e_1，$\cdots$，e_α 都合格，且 e^α 从 $(f,\ \varepsilon^{\alpha-1})$ 依数组调整术中诸规则获得（$\varepsilon^0=\varepsilon$）。特别有，数组 $\varepsilon^L=(\varepsilon_{ij}^L)$ 不仅满足 $(\mathrm{I})_f$ 且同时满足 $(\mathrm{II})_f$，因而从 ε^L 可依第三部分第 5 节中线图嵌入第一基本定理具体作出整个线图 G 的嵌入：

$$g: G\subset\text{平面}。$$

五、（平面性）线图嵌入的分类

1．树形嵌入的扩充

问题 在第二部分第5节的嵌树术中，解决了树形嵌入的分类问题。在第三部分第3节扩充术与第三部分第5节，第四部分第7节的线图嵌入术中，解决了一个线图 G 的最大树形 T 应如何嵌入才有可能扩充到整个线图 G 的嵌入的问题。但是，当已有嵌入 $g: T\subset$ 平面并知 g 可扩充为 G 的嵌入时，这种扩充的方法仍可不止一种。在第三部分第1节交截数以及第三部分第6节诸节中，已有例说明此事，见图12与26。这些事例导致下面的问题：

设 $g: T\subset$ 平面可扩充为 G 的嵌入，问如何区别并完全定出各种不同的扩充类型。

剖析 已有嵌入 $g: T\subset$ 平面而要扩充到某一外棱 e_α，显然有两种可能方法，视嵌入后 $e_\alpha\cup P_\alpha=C_\alpha$ 含树根于其内部或外部而定。为此我们对每一外棱 e_α 引入一模2数 u_α 以表示 e_α 的嵌入方式：

$$u_\alpha=\begin{cases}0,\ e_\alpha\text{ 嵌入后使树根在 }C_\alpha\text{ 之外，}\\1,\ e_\alpha\text{ 嵌入后使树根在 }C_\alpha\text{ 之内。}\end{cases}$$

每一个扩充到 G 的嵌入都有一确定的数组（u_α）与之对应，但并不是任意一个数组都有一相应的扩充。例如在图26的例中（树根位于 t_1，t_3 之间，未在图中标出），（u_1，u_2，u_3）所能取的值只能是

（1，0，0），（1，1，0），（1，1，1），（0，0，0）

4种，而不能取另外的4种。其原因是：当一外棱 e_α 已嵌入而再要嵌入另一外棱 e_β 时，其方式可能有两种，也可能只有一种。例如图26中，若 e_1 已嵌入使 $u_1=1$，则再嵌入 e_2 时有

两种方式使 $u_2=0$ 或 1，其次在 e_1，e_2 已嵌入使 $u_1=1$，$u_2=1$ 再嵌入 e_3 时可有两种方式使 $u_3=0$ 或 1。但若 e_1，e_2 已嵌入使 $u_1=1$，$u_2=0$，则再嵌入 e_3 时，只能再有一种方式，即 $u_3=0$。

这说明对任两外棱 e_α，e_β 来说，u_α，u_β 的取值是不能任意的，数组（u_α）的取值更不能任意。由此可知，为解决原来的问题，首先应求得对任两外棱 e_α，e_β 来说，数 u_α 与 u_β 之间的制约关系。

为此，记 e_α 的两端为 a，a'，e_β 的两端为 b，b'。由于已知 $g:T\subset$平面可扩充为 G 的嵌入，故由第三部分第 3 节扩充术知

$$g_{\alpha\beta}=0。$$

因而若 e_α 已嵌入平面，则 b，b'将同在 $C_\alpha=e_\alpha\cup P_\alpha$ 之内或同在其外。同样，若 e_β 已嵌入平面，则 a，a'也将同在 $C_\beta=e_\beta\cup P_\beta$ 之内或同在其外。又

$$g_{a\beta}=0 \text{ 或 } 1$$

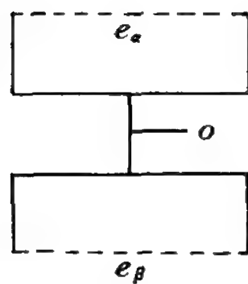
视 a 与树根在 g 下位于 P_β 同侧或异侧而定。在 g 扩充至 e_α 后，则视是否同在 C_β 之内外或一在内而另一在外而定。故其值与嵌入 $g:T\subset$平面有关而与 g 的扩充无关。对 g_{ba} 亦然。由这些几何意义可知，应有

$$g_{a\beta}g_{ba}=0,$$

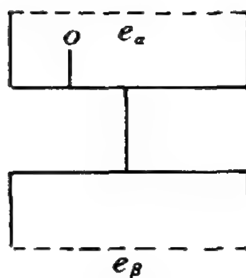
即 $g_{a\beta}$，g_{ba}不能同时为 1。

今任将 $g:T\subset$平面扩充为 G 的一个嵌入，则就 C_α，C_β 以及根的相互内外位置关系而言可概括为图 32 中的 9 种类型(实际嵌入都可归入这些类型之中，只是某些点线可能重合)。

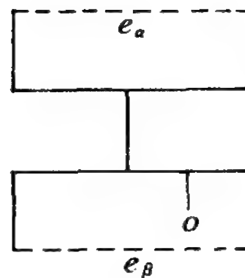
对于这些不同情形，u_α 等所取的值可列表如下：



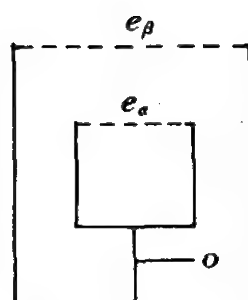
(1)



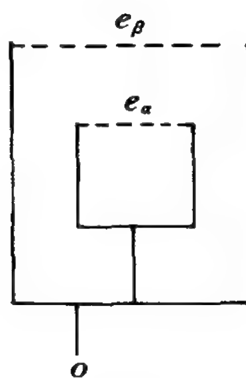
(2)



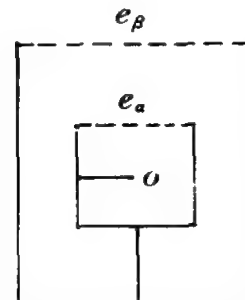
(3)



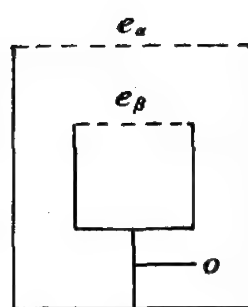
(4)



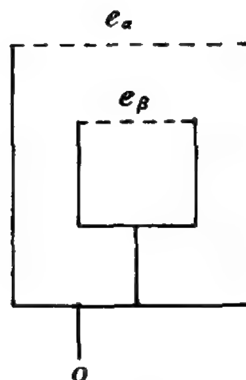
(5)



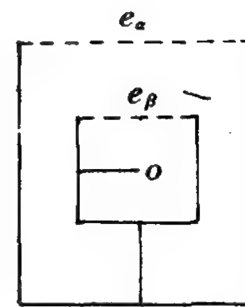
(6)



(7)



(8)



(9)

图 32

情　　形	u_α	u_β	$g_{\alpha\beta}$	g_{ba}
1	0	0	0	0
2	1	0	0	1
3	0	1	1	0
4	0	1	0	0
5	0	0	1	0
6	1	1	0	1
7	1	0	0	0
8	0	0	0	1
9	1	1	1	0

数组

$$(u_\alpha,\ u_\beta,\ g_{\alpha\beta},\ g_{ba})$$

所能取的不同组值为 $2^4=16$。但由于

$$g_{\alpha\beta}g_{ba}=0,$$

实际上所能取的组值只有 12 个。在此 12 个组值中，只有上表所示的 9 个解为某一嵌入所实现，其他 3 组则不可能。我们现在的目的，在寻觅一个 u_α，u_β，$g_{\alpha\beta}$，g_{ba}间的代数式 Y，使对上表中 9 种情形都有 $Y=0$，而对其他三种情形则有 $Y=1$。这样 $Y=0$ 将为 e_α，e_β 各种可能嵌入的制约关系。

显然 Y 不能取为 u_α，u_β，$g_{\alpha\beta}$，g_{ba} 间的一次式。为对符合要求的各种二次式进行尝试，先列出一表如下：

由上表可见，一个合适的代数式是

$$Y=u_\alpha u_\beta+u_\alpha g_{\alpha\beta}+u_\beta g_{ba}。$$

因此，数组（u_α）相应于一真正扩充到 G 的嵌入的必要条件是对任两外棱 e_α，e_β 都有

$$u_\alpha u_\beta + u_\alpha q_{\alpha\beta} + u_\beta{}^{3}{}_{ba} = 0。$$

容易验证，这个条件也是充分的。我们把这一结论表达于下，并于其后的验证中证实这一结论的真实性。

情形		数组 $(u_\alpha,\ u_\beta,\ g_{\alpha\beta},\ g_{ba})$	$u_\alpha u_\beta$	$u_\alpha g_{\alpha\beta}$	$u_\beta g_{ba}$
可能情形	1	(0, 0, 0, 0)	0	0	0
	2	(1, 0, 0, 1)	0	0	0
	3	(0, 1, 1, 0)	0	0	0
	4	(0, 1, 0, 0)	0	0	0
	5	(0, 0, 1, 0)	0	0	0
	6	(1, 1, 0, 1)	1	0	1
	7	(1, 0, 0, 0)	0	0	0
	8	(0, 0, 0, 1)	0	0	0
	9	(1, 1, 1, 0)	1	1	0
不可能情形	10	(1, 0, 1, 0)	0	1	0
	11	(0, 1, 0, 1)	0	0	1
	12	(1, 1, 0, 0)	1	0	0

树形嵌入扩充分类术　设 $g: T \subset$ 平面，已知 g 可扩充为 G 的嵌入。对每一外棱 e_α，引入一（模 2）未知数 y_α，则扩充为 G 的嵌入的不同方式与下述（模 2）方程组（a 是 e_α 的任一端，b 是 e_β 的任一端）：

$$(Y)_g \qquad\qquad Y_{\alpha\beta}(g) = 0$$

的不同(模 2)解答$(y_\alpha) = (u_\alpha)$相对应，其中

$$Y_{\alpha\beta}(g) = y_\alpha y_\beta + y_\alpha g_{\alpha\beta} + y_\beta g_{ba}。$$

验证　在“剖析”中已说明每一扩充到 G 的嵌入必相应于$(Y)_g$的一组解答，反之，设

$$(y_\alpha) = (u_\alpha)$$

是$(Y)_g$的一组解答，试用归纳法验证必有一扩充到G的嵌入与之相应如下。

将G的外棱任意排成一次序$e_1, e_2, \cdots, e_L$，设已将$g: T\subset$平面扩充为一嵌入

$$\bar{g}: T\cup e_1\cup\cdots\cup e_k\subset \text{平面},$$

使

$$u_\beta = \text{次}\ o\bar{g}(C_\beta), \beta=1, \cdots, k。$$

试考虑

$$e_\alpha = e_{k+1}。$$

按$(Y)_g$，有

$$u_\alpha(u_\beta + g_{\alpha\beta}) = q_{b a}u_\beta, \beta=1, \cdots, k。$$

若对每一$\beta=1, \cdots, k$，都有

$$u_\beta + g_{\alpha\beta} = 0,$$

亦即

$$\text{次}_a\bar{g}(C_\beta)=0。$$

由此知e_α的任一端a，因而两端a, a'都在诸$\bar{g}(C_\beta)$之外。因而可有两种方式扩充为e_α的嵌入。我们可选取$\bar{g}(e_\alpha)$，使

$$\text{次}_Q\bar{g}(C_\alpha) = \text{给定的}\ u_\alpha。$$

其次，设在$\beta=1, \cdots, k$中至少有一，设为β_0，使

$$u_{\beta_0} + g_{\alpha\beta_0} = 1,$$

则仍按$(Y)_g$，有

$$u_\alpha = g_{b a}u_{\beta_0}。$$

由于已给u_{β_0}而$g_{b a}$由$g: T\subset$平面完全确定，故由上式知，u_α已由g及其扩充$\bar{g}: T\cup e_1\cup\cdots\cup e_k\subset$平面唯一确定。因已知$g_{\alpha\beta}=0, \beta=1, \cdots, k$，故知$\bar{g}$必可扩充至$e_\alpha$，且知扩充所得的数

$$次\ o\,\bar{g}(C_\alpha)$$

只能是上面唯一确定的 u_α。

因之,由归纳法知,必有扩充

$$\bar{g}: G\subset 平面,$$

使诸外棱 e_α 所定数组

$$(次\ o\,\bar{g}(C_\alpha))$$

即为方程组$(Y)_g$ 预先给定的解答(u_α)。至此验证完毕。

2.（平面性）线图嵌入的分类（第四基本定理）

问题 上面树形嵌入扩充分类术是对一确定且已知可扩充到 G 的嵌入 $g: T\subset$平面而言。试对任意已知可扩充到 G 的嵌入 $f: T\subset$平面作出相应的结论。

剖析 对于任一嵌入 $f: T\subset$平面，由第三部分第 5 节与第四部分第 5 节中的线图嵌入术（第一，第三基本定理）知，f 能否改变为 $g: T\subset$平面，使 g 可扩充为 G 的嵌入系由方程组

$$(\mathrm{I})_f \qquad x_{\alpha\beta} = f_{\alpha\beta},$$

$$(\mathrm{II})_f \qquad Q_{ijk}(f, x) = 1.$$

所定,且这样的嵌入 $g: T\subset$平面与方程组$(\mathrm{I})_f$,$(\mathrm{II})_f$ 的解答

$$(x_{ij}) = (\varepsilon_{ij})$$

相对应,使

$$g_{ij} = f_{ij} + \varepsilon_{ij}。$$

对于一个确定的、可扩充至 G 的嵌入 $g: T\subset$平面,亦即对于一组确定的$(\mathrm{I})_f$,$(\mathrm{II})_f$ 的解答$(x_{ij}) = (\varepsilon_{ij})$,$g$ 扩充到整个 G 的可能的嵌入依树形嵌入扩充分类术又与方程组

$$(Y)_g \qquad Y_{\alpha\beta}(g) = 0$$

的解答相对应,而 $Y_{\alpha\beta}(g)$中

$$q_{\alpha} = f_{\alpha\beta} + \varepsilon_{\alpha\beta},$$

$$q_{ba} = f_{ba} + \varepsilon_{ba}。$$

由此得下述结论：

线图嵌入分类术（第四基本定理） 任作 $f: T \subset$ 平面，则 G 可嵌入平面的类型与下方程组$(\mathrm{I})_f$，$(\mathrm{II})_f$，$(\mathrm{III})_f$ 的解答数相对应：

$(\mathrm{I})_f$ $\quad x_{\alpha\beta} = f_{\alpha\beta}$。

$(\mathrm{II})_f$ $\quad (x_{ij} + f_{ij})(x_{ik} + f_{ik}) + (x_{jk} + f_{jk})(x_{ji} + f_{ji}) + (x_{ki} + f_{ki})(x_{kj} + f_{kj}) = 1$。

$(\mathrm{III})_f$ $\quad y_{\alpha}y_{\beta} + (x_{\alpha\beta} + f_{\alpha\beta})y_{\alpha} + (x_{b\alpha} + f_{b\alpha})y_{\beta} = 0$。

总　　结

在上面五部分中，对布线诸问题的解决作了详细说明。现将所用方法与所得结论重新概述如下。

首先在已给线图 G 中任选一最大树 T 并添加一树根 O。G 限制如第一部分第 1 节中所说，且 G，T 已改造如第二部分第 2 节中所示。

将 T 任意嵌入平面，嵌入方式记为 f，即作

$$f: T \subset \text{平面},$$

于是对任两以同一顶点为前端的树棱 t_i，t_j，依 f 下 t_i，t_j 旋向的顺反可定一模 2 数

$$f_{ij} = f_{ji} + 1。$$

又任意扩充 f 为一相对于 T 的浸入

$$\bar{f}: G \underset{T}{\propto\subset} \text{平面},$$

则对每两外棱 e_{α}，e_{β}，由 $\bar{f}$ 下 e_{α}，e_{β} 的交截情况可定一与扩充 $\bar{f}$ 无关的模 2 数

$$f_{\alpha\beta}=f_{\beta\alpha}。$$

在数组$(f_{\alpha\beta})$与(f_{ij})间有基本关系式

$$f_{\alpha\beta}=f_{ab}+f_{ab'}+f_{a'b}+f_{a'b'},$$

此处 $e_\alpha=\widehat{aa'}$, $e_\beta=\widehat{bb'}$，而 f_{ab}等由 f_{ij}等确定如第三部分第 4 节中所示。

原来 T 的嵌入 f 一般不可能扩充到整个 G，而需先将 f 变为另一 T 的嵌入

$$g: T\subset 平面$$

始有此可能。又在能扩充为 G 的嵌入时，每一外棱 e_α在嵌入后与树形是否共同包含树根 O 于其内部又有两种可能情况。为此，我们引入两个模 2 未知数组

$$(x_{ij})与(y_\alpha),$$

其中 $x_{ij}=x_{ji}$描述以同一顶点为前端的两树棱 t_i, t_j在 f 下的旋向是否需要改变，而 y_α则描述外棱 e_α嵌入后是否应与树形共同包含根 O 于其内部。

我们确定了三组方程

$(\mathrm{I})_f$ $\qquad x_{\alpha\beta}=f_{\alpha\beta}。$

$(\mathrm{II})_f$ $\qquad (x_{ij}+f_{ij})(x_{ik}+f_{ik})+(x_{jk}+f_{jk})(x_{ji}+f_{ji})$
$\qquad\qquad +(x_{ki}+f_{ki})(x_{kj}+f_{kj})=1。$

$(\mathrm{III})_f$ $\qquad y_\alpha y_\beta+(x_{\alpha\beta}+f_{\alpha\beta})y_\alpha+(x_{ba}+f_{ba})y_\beta=0。$

这些方程的几何意义分别如下：

方程$(\mathrm{II})_f$ 代表将f: $T\subset$平面改为另一 T 的嵌入时旋向的改变值(x_{ij})必须满足的制约关系。方程$(\mathrm{I})_f$ 代表将f 依(x_{ij})改变为另一 T 的嵌入后，可扩充至外棱 e_α, e_β，使在此嵌入下不相遇的条件。方程$(\mathrm{III})_f$ 则代表 f 已依(x_{ij})改变为另一嵌入后，为扩充至 G 诸外棱在嵌入下必须遵守的相互缠绕的制约关系。

我们获得的结果如下：

(1) G 可嵌入的充要条件是方程组 $(\mathrm{I})_f$ 有解——第二基本定理。

(2) 方程组 $(\mathrm{I})_f$ 有解时，联立的方程组 $(\mathrm{I})_f$，$(\mathrm{II})_f$ 也有解，且从 $(\mathrm{I})_f$ 的解可依确定步骤获得 $(\mathrm{I})_f$，$(\mathrm{II})_f$ 的解——第三基本定理。

(3) 可以扩充到整个 G 的不同嵌入 $g: T\subset$ 平面恰与 $(\mathrm{I})_f$，$(\mathrm{II})_f$ 的不同解答 (x_{ij}) 相对应——第一基本定理。

(4) G 的所有不同嵌入

$$g: G\subset \text{平面},$$

恰与 $(\mathrm{I})_f$，$(\mathrm{II})_f$，$(\mathrm{III})_f$ 联立方程组的不同解答

$$(x_{ij},\ y_\alpha)$$

相对应——第四基本定理。

以上 (1) — (4) 解答了图 7 中的全部问题，只是机器化程序的编试尚待进行。但这些结果都通过选取一最大树 T 以及任择一嵌入 $f: T\subset$ 平面而获得，而 T 与 f 的选取都带有极大的任意性。为此，可提出下述问题：考虑一种由 G 本身而不依赖于 T 与 f 的选取的解决方法。这个方法可由本书的示嵌类理论提供。按从 G 可作一、二维模 2 示嵌类 $\phi^2(G)$。我们的一般理论指出

$$\boxed{G \text{ 可嵌入平面} \Leftrightarrow \phi^2(G)=0}$$

这一结果称为线图嵌入的第五基本定理。若选择一最大树 T 与任意嵌入 $f: T\subset$ 平面，容易验明

$$\boxed{\phi^2(G)=0 \Leftrightarrow (\mathrm{I})_f \text{ 有解}}$$

由此即得上面的结果 (1)。由于这些结论是纯理论性的，与本附录宗旨不尽相符，故论证一概从略。

附记 (1977 年 12 月)。方程组 $(\mathrm{I})_f$ 的每一个方程一

般说来可以有 4 个含未知数的项。最近数学所的刘彦佩同志证明，若适当选择最大树，可使（Ⅰ）$_f$ 的每一方程都祇含 2 个有未知数的项。这使判断（Ⅰ）$_f$ 有解与否也即 G 能否嵌入平面与否变得极为简单。这是本理论的一个重要进展。刘彦佩同志的论文《模 2 规划与平面嵌入》将发表于应用数学学报。

集成电路设计中的一个数学问题*

一、引　　言

集成电路的设计牵涉到线性图的平面性即是否可在平面中拓扑地实现的问题，这个问题的各个方面可用图 1 来概括：

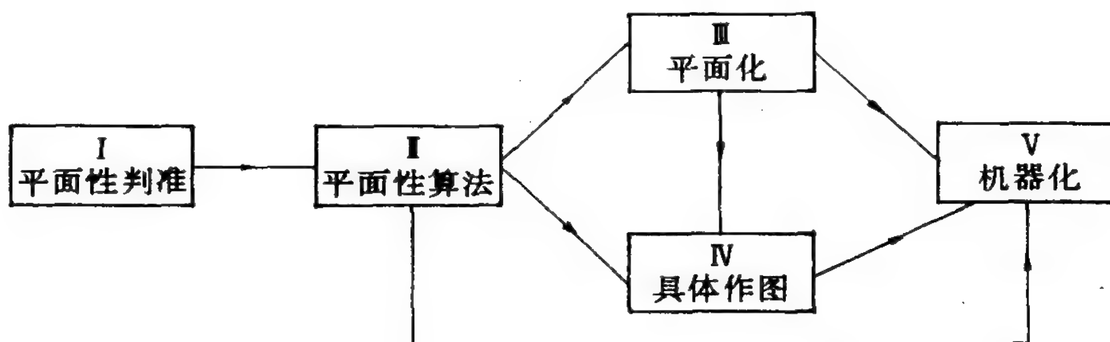


图 1

I—V 的几个方面可说明如下：

I．**平面性判准**——给出一个线性图是否有平面性的判准。这样的判准早在 30 年代就已给出，如 Kuratowski[4]，Whitney[5]，MacLane[5]等。其中 Kuratowski 判准说，线性图 G 有平面性的充要条件是，G 不含有任何下面两类子图形之一：

但这一类判准在实用上有困难，例如图 3 是非平面性的，有一 Kuratowski 子图形如粗黑线所示，但并不容易找出：

---

* 本文摘自《数学认识与实践》，1973 年。

Ⅱ．平面性算法——给出一个算法，依据它能依照确定步骤来判断一个线性图是否有平面性。上面所提到的那些判准都不足以从此导出一个算法来。这样的算法直到近几年才出现，见 Auslander-Parter [1], 与 Bader[2]。

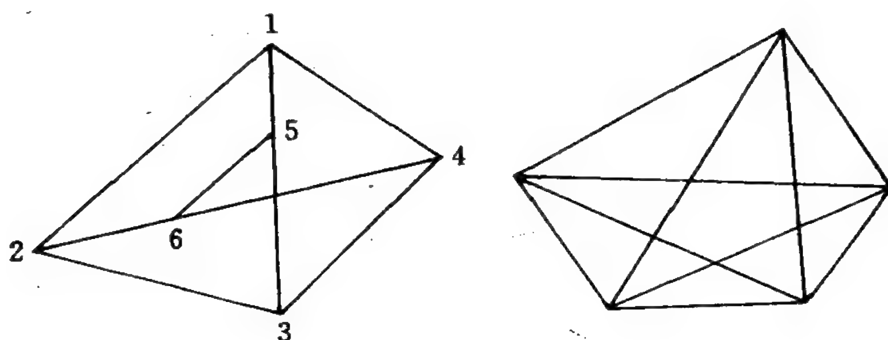


图 2

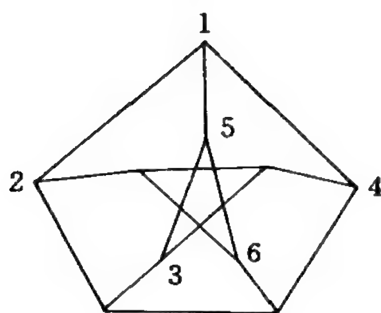


图 3

Ⅲ．平面化——原来的线性图是非平面性的，但在除去某些棱后余下的线性图可以是平面性的。这种手续可叫做线性图的平面化。

Ⅳ．具体作图——即使线性图 G 有平面性或已平面化，也并不等于说 G 已在平面上画出，因之需要有一个将 G 在平面上具体画出的方法。

Ⅴ．机器化——由于所考虑到的线性图往往是很复杂的，即使Ⅱ—Ⅳ诸问题在原则上已解决也难以用手算来付诸实施。因之需要程序化，使Ⅱ—Ⅳ各问题都能在电子计算机上解决。

国外对这些问题的处理情况如下：

Auslander-Parter 与 Bader（两文内容实质上是相同的）给出了一个分解原来线性图为更小的图形并重复进行的方法来解决问题Ⅱ—Ⅳ，Fisher-Wang［3］更依据它来机器化，但是他们所给出的只是一个具体进行的方法而缺少理论的提炼，而 Kuratowski 等则虽有理论却不能实际使用。

我们对这些问题的处理情况如下：

对问题Ⅰ，我们在 1954 年以来所创立的示嵌类理论早已提供了一个简单的判准。这个判准容易化为一组模 2 系数的线性方程组的可解性问题，在这组方程有解时，根据它的解即可具体作图。这给出了问题Ⅱ—Ⅳ的解答，依据它来机器化乃是一个纯粹程序设计的问题，并没有什么实质性的困难。

二、线性图的平面性与示嵌类

设 G 是一线性图，G 的顶点的集合将记作 G^0，棱的集合记作 G^1。我们将作下述假定，这对 G 的平面性的考虑不产生影响：

（1）没有两端是同一顶点的棱。

（2）连接 G 中任两顶点的棱，至多只有一条。

（3）G 是连通的。

记 G^1 中的棱为 e_1，…，e_n，指数集 $\{1, \cdots, n\}$ 记作 I。对于一切无序指数偶 (i, j)，$i \neq j$，i，$j \in I$，我们将分成两部分 D 与 N。如果 e_i，e_j 在 G 中无公共顶点，则置 $(i, j) \in D$，否则置 $(i, j) \in N$。

对称约化积

对 G 可作一二维复形 G^*，称之为 G 的二重对称约化积者如次，G^* 的二维胞腔对应于一切 $(i, j) \in D$，记作 $e_i{}^* e_j$

$=e_j{}^* e_i$。G^* 的一维胞腔对应于 G 中一切顶点 v 与棱 e_i 的偶合，而 v 非 e_i 的顶点，记之为 $v^* e_i = e_i{}^* v$。G^* 的 0 维胞腔对应于 G 中一切无序的顶点偶（v，w），而 $v \neq w$，记之为 $v^* w = w^* v$。如果 e_j 的顶点是 v_{i1}，v_{i2}，e_j 的顶点是 v_{j1}，v_{j2}，则 G^* 中的（模 2）边界关系将定义为

$$\begin{cases}\partial(e_i{}^* e_j) = v_{i1}{}^* e_j + v_{i2}{}^* e_j + v_{j1}{}^* e_i + v_{j2}{}^* e_i, \\ \partial(v^* e_i) = v^* v_{i1} + v^* v_{i2}, \\ \partial(v^* w_i) = 0。\end{cases}$$

如果引入以下规定：

$$\begin{cases} e_i{}^* e_j = 0, (i,j) \in N \text{ 时}; v^* e_i = 0, v \text{ 为 } e_i \text{ 端点时}; v^* v = 0, \\ \sum\limits_{i,j} a_{ij} e_i{}^* e_j = \sum\limits_j (\sum\limits_i a_{ij} e_i) * e_j = \sum\limits_i (e_i{}^* \sum\limits_j a_{ij} e_j), \end{cases}$$

余类推，则以上边界关系亦可表作（$e_i{}^* e_j, v^* e_i, v^* w \in G^*$）

$$\begin{cases}\partial(e_i{}^* e_j) = \partial e_i{}^* e_j + e_i{}^* \partial e_j, \\ \partial(v^* e_i) = v^* \partial e_i, \\ \partial(v^* w) = 0,\end{cases}$$

或对偶地说，

$$\begin{cases}\delta(v^* w) = \delta v^* w + v^* \delta w, \\ \delta(v^* e_i) = \delta v^* e_i, \\ \delta(e_i{}^* e_j) = 0。\end{cases}$$

浸入与嵌入

以 R^2 表平面。映象 $f: G \to R^2$ 将称为一个浸入，记为 $f: G \subset R^2$，如果以下条件满足：

（1）对每一棱 $e_i \in G^1$，f/e_i 是一拓扑映象。

（2）对每一棱 $e_i \in G^1$，$f(e_i)$ 除原来的端点外，不再经过其他顶点的象。

（3）对每两棱 $e_i \neq e_j \in G^1$，$f(e_i)$，$f(e_j)$ 除可能有的公共顶

点外，至多“简单相交”于有限个内点，即每一交点处的相交指数为 1(模 2)。

如果浸入 f 对每两棱 $e_i \neq e_j \in G^1$，$f(e_i)$，$f(e_j)$ 至多在可能有的公共顶点处相遇，则 f 是 G 的一个拓扑映象。此时 f 将称为一个嵌入，记作 $f: G \subset R^2$。如果有这样的嵌入存在，则 G 称为一平面图形，或具有平面性，记作 $G \subset R^2$。

示嵌链

所有的链，同调等都将是模 2 的，在以下不再说明。

· 设 $f: G \propto R^2$ 是一个浸入。对 f 引入一 G^* 中的二维上闭链 $\varphi_f = \varphi$ 如下。对任意 $(i,j) \in D$，命

$$\varphi_{ij} = I_f(e_i, e_j),$$

即 $f(e_i)$、$f(e_j)$ 的相交指数。置

$$\varphi = \sum_{(i,j) \in D} \varphi_{ij} e_i{}^* e_j。$$

显然 φ 是上闭链，我们将称 φ 为浸入 f 的示嵌链。φ 之不等于 0 可视为 f 异于嵌入的一个测度。

定理 对任意浸入 f，相应的示嵌链 φ 恒属于同一个上同调类 Φ。

证 设 $f, g: G \propto R^2$，相应的示嵌链各为 $\varphi = \sum \varphi_{ij} e_i{}^* e_j$ 与 $\psi = \sum \psi_{ij} e_i{}^* e_j$。需证 $\varphi \sim \psi$。

为此，先设对任意顶点 v，$f(v) = g(v)$，且有一 $e_k \in G^1$，在 $e_i \neq e_k$ 时，$f \equiv g / e_i$，而 $f(e_k)$，$g(e_k)$ 除公共端点外不相遇。于是 $f(e_k)$，$g(e_k)$ 构成一 R^2 中的简单闭曲线 C。设在 f 与 g 下象在 C 内部的顶点为 $v_{k1}, \cdots, v_{ks}$。试证

$$\varphi = \psi + \sum_{r=1}^{s} \delta v_{k_r} * e_k。 \quad (*)$$

为此，先设 $i \neq k$，$j \neq k$，而 $(i,j) \in D$，则显然有 $\varphi_{ij} = \psi_{ij}$，而 $(*)$ 式左右两边 $e_i * e_j$ 的系数相等。今设 $(k, l) \in D$，而 e_l 两端点在 f, g 下的象同在 C 的内部。则此两端点将为 v_{kp} 与 v_{kq}。于

是在(*)式右边的$\sum$中$e_k * e_l$将出现两次，而其和为0。其次$f(e_l)=g(e_l)$与C将简单相交于偶数次，因而$\varphi_{kl}=\psi_{kl}$。故此时两边$e_k{}^* e_l$的系数仍相等。当两端点在f,g下的象同在C外部时亦如此。最后设两端点的象一在C内而一在C外，在内者设为v_{kp}。则(*)式右边$\sum$中$e_k * e_l$的项将出现一次。又此时$f(e_l)=g(e_l)$将与C简单相交于奇数个点，因而$\varphi_{kl}=\psi_{kl}+1$。故两边$e_k{}^* e_l$的系数仍将相等。因而(*)式成立而有$\varphi\sim\psi$。

在上情形中$f(e_k),g(e_k)$除公共顶点外不相遇的条件容易除去。盖可取$h:G\subset R^2$使$e_i\neq e_k$时，$h\equiv f\equiv g/e_i$，而$h(e_k)$，$f(e_k)$以及$h(e_k),g(e_k)$各满足上述条件。命h的示嵌链为θ，则将有$\theta\sim\varphi$与$\theta\sim\psi$，因而$\varphi\sim\psi$。

次设对任意顶点v有$f(v)=g(v)$，而对任意两棱$e_k,e_l\in G_1$(k亦可$=l$)，$f(e_k)$与$g(e_l)$除可能的公共端点外，只简单相交于有限个内点，即在每点的相交指数为1(模2)。今作$h_0,\cdots,h_{n-1}:G\subset R^2$如下。$h_k:G\subset R^2$，使$h_k$在$e_1,\cdots,e_k$上与$f$相同而在$e_{k+1},\cdots,e_n$上与$g$相同。则$h_0\equiv g$而$h_n\equiv f$。记$h_k$的相应示嵌链为$\varphi_k$。由已证明的情形，应有$\varphi_k\sim\varphi_{k+1}$，$k=0,1,\cdots,n-1$。故得$\varphi_0\sim\varphi_n$，即$\psi\sim\varphi$。

若对任意顶点v，仍有$f(v)=g(v)$。则总可作一浸入$h:G\subset R^2$，使对任意顶点v，有$h(v)=f(v)=g(v)$，而对任意两棱$e_k,e_l\in G^1$(k亦可$=l$)，$f(e_k)$与$h(e_l)$以及$g(e_k)$与$h(e_l)$除可能的公共端点外，只简单相交于有限个内点，记h的相应示嵌链为θ，则应有$\varphi\sim\theta$，$\psi\sim\theta$，故仍有$\varphi\sim\psi$。

最后，设f,g任意。命h为R^2到R^2上的一个拓扑变换，使对任意顶点$v\in G^0$，有$hf(v)=g(v)$。记hf的相应示嵌链为θ，则依前述情形有$\theta\sim\psi$。另一面因h为拓扑变换，故显然有$\theta\sim\varphi$。因之仍有$\varphi\sim\psi$。

至此定理证毕。

定义 定理中所确定的上同调类 Φ 将称为 G 的示嵌类。

如果 G 有平面性，而 $f: G \subset R^2$ 是一嵌入，则相应于 f 的示嵌链，显然 0。因而可得下述。

嵌入定理(必要部分) G 有平面性的一个必要条件是：G 的示嵌类 $\Phi = 0$，或对任意浸入 $f: G \propto R^2$，相应的示嵌链 $\varphi \sim 0$。

(注：这个定理中的条件不仅是必要的，而且是充分的。但这一点将在五中证明。)

三、G 的平面性判准——基本定理与基本方程组

为了根据嵌入定理以确定 G 是否具有平面性，试记 G 中顶点的集合为 $\{v_1, \cdots, v_m\} = G^0$。任作一浸入 $f: G \propto R^2$，记其相应的示嵌链为

$$\varphi = \sum \varphi_{ij} e_i{}^* e_j。$$

于是 $G \subset R^2$ 的一个必要条件是 $\varphi \sim 0$ 或有 $x\lambda_i$ 存在，使

$$\varphi = \sum x_{\lambda 1} \delta(v_\lambda * e_i),$$

这里的 $\sum$ 展开于一切使 v_λ 非 e_i 端点的指数偶(λ, i)上。比较两边 $e_i * e_j$ 的系数，即得一组以 $x_{\lambda i}$ 为未知数的模 2 线性方程组。这组方程可解是 G 有平面性的必要条件，以后也将证明这些条件是充分的。

这个判准的原理虽然简单，但由于未知数的个数接近于 mn，方程的个数接近于 n^2，因之当图 G 比较复杂，m，n 相当大时，这个判准事实上并不实用，即使用了电子计算机也是如此。因此有必要简化上述方程组，使之即使在 m，n 很大时，也可以通过电子计算机来实现。这是以下的目的所在。

最大树

为了这一目的，试在 G 中取一树 T，通过 G 的所有顶点。这样的树 T 将称为 G 的一个最大树。T 的任一棱将称为一个节段，顶点也称节点。如果 T 中在节点 v_λ 处只有一个节段，则 v_λ 也称为节梢。

对 T 的任一节段 e_i，T—内部(e_i)分成两个分支，记为 C'_i，C_i''。命 $\sum_i$ 为所有 $e_a \in G^1 - T^1$ 的集合，这里 e_a 的一端在 C_i' 中，而另一端在 C_i''中。又对任一棱 $e_a \in G^1 - T^1$，在 T 中将有唯一的一条通路从 e_a 的一端到 e_a 的另一端，记之为 $\mathscr{P}_a$。置

$$E_i = \sum_{e_a \in \Sigma_i} e_a, (e_i \in T^1),$$

$$P_a = \sum_{e_i \in \mathscr{P}_a} e_i, (e_a \in C^1 - T^1),$$

这里 E_i, P_a 都是 G 中的(模 2)上链。于是显然有

$$e_a \in \sum{}_i \Leftrightarrow e_i \in \mathscr{P}_a。$$

引理 1 对于最大树 T 而言，G 的一维上边缘群，有一组基底由以下各上边缘所构成：

$$e_i + E_i, e_i \in T。$$

证 对 $e_i \in T$，命 $A_i'(A_i'')$为 $C_i'(C_i'')$中一切顶点的集合，则易见

$$e_i + E_i = \sum_{v' \in A_i'} \delta v' = \sum_{v'' \in A_i''} \delta v''。\qquad (1)'$$

故每一 $e_i + E_i$ 为一上边缘，且不同的 $e_i + E_i$ 恰含有一不同的 $e_i \in T$，故这些 $e_i + E_i$ 是线性独立的。

另一面，对任一 G 的顶点 v，设 e_j，e_{j1}，…，e_r 是 T 中在 v 处的节段全体。命 C_j'为 T－内部（e_j）中含有 v 的分支，而 C'_{js}为 T－内部（e_{js}）中不含 v 的那一分支。则由（1）式有

$$\delta_v = (e_j + E_j) + \sum_{s=1}^{r} (e_{js} + E_{js})。$$

故任一上边缘可表为形如 $e_i + E_i$ 的上边缘之和。由此引理得

证。

（注：上述引理事实上是已知的，在电网络理论中 e_i+E_i 是一割集（cut set），参阅 S. Seshu-N. Balanian, Linear Network Analysis, 1959, 页 70—71）

引理 2 G^* 的二维上边缘群有以下一组生成元：

$$\begin{cases}(e_i+E_i)*(e_j+E_j),e_i,e_j\in T^1,(i,j)\in D\cup N;\\(e_i+E_i)*e_\alpha,e_i\in T^1,e_\alpha\in G^1-T^1,(i,\alpha)\in D\cup N。\end{cases}$$

证 命 v 为 G 的任一顶点。e 为 G 的任一棱，而 v 非 e 的顶点，则 G^* 的二维上边缘群由诸 $\delta(v*e)=\delta v*e$ 所生成。因 G 的一维上链群有一组基由诸 $e_j\in T^1$ 与 $e_\alpha\in G^1-T^1$ 所构成，故也有一组基由诸 e_j+E_j 与 e_α 所构成。另一面由引理 1，δv 又可表为诸 e_i+E_i，$i\in T^1$ 的线性和，故得本引理。

定义 对最大树 T 而言，一个浸入 $f:G\subset R^2$ 将称为一个 T－浸入，如果（1）$f:T\subset R^2$，（2）对任意 $e_i\in T^1$ 与 $e_\alpha\in G^1-T^1$，$f(e_i)$ 与 $f(e_\alpha)$ 至多在公共端点处相遇。

引理 3 设 $\tilde{T}$ 是 T 在 G 中的一个邻域，则一个 T－浸入 f 的示嵌链由 $f/\tilde{T}$ 所完全决定。换言之，若 f，g 是两个 T－浸入，而 $f/\tilde{T}\equiv g/\tilde{T}$，则 $\varphi_f=\varphi_g$。

证 若 e_i，$e_j\in T^1$，$e_\alpha\in G^1-T^1$，而 (i,j)，$(i,\alpha)\in D$，则 $\varphi_f(e_i*e_j)=\varphi_g(e_i*e_j)=0$，$\varphi_f(e_i*e_\alpha)=\varphi_g(e_i*e_\alpha)=0$。今设 e_α，$e_\beta\in G^1-T^1$，$(\alpha,\beta)\in D$。于是 $f(e_\alpha),f(g_\alpha)$ 构成一简单闭曲线 C。记 e_β 的两端为 a_1，a_2，则 $f(a_1)=g(a_1)$，$f(a_2)=g(a_2)$。在 e_β 内部取两点 a_1'，$a_2'\in\tilde{T}$ 充分接近于 a_1，a_2。则视 a_1'，a_2' 同在 C 之内或外或一在 C 内一在 C 外，相交系数 $I(f(e_\beta),f(e_\alpha))$ 与 $I(g(e_\beta),f(e_\alpha))$ 将同为 0 或同为 1。同样，相交系数 $I(g(e_\beta),f(e_\alpha))$ 与 $I(g(e_\beta),g(e_\alpha))$ 亦同为 0 或 1。故

$$I(f(e_\beta), f(e_\alpha)) = I(g(e_\beta), g(e_\alpha))$$

或

$$\varphi_f(e_\alpha * e_\beta) = \varphi_g(e_\alpha * e_\beta)。$$

因而 $\varphi_f = \varphi_g$，如所欲证。

基本定理 I （必要部分）　G 有平面性的一个必要条件是，对最大树 T 而言，下述方程组有解：

$$\begin{cases} \sum_0 x_{ij} + \sum_1 x_{i\alpha} + \sum_2 x_{i\beta} = \varphi_{\alpha\beta}, \\ (\alpha,\ \beta) \in D,\ e_\alpha,\ e_\beta \in G^1 - T^1。 \end{cases} \tag{1}$$

其中

$\varphi = \sum \varphi_{\alpha\beta} e_{\alpha\alpha} e_\beta$ 为相应于某一 T - 浸入的示嵌链。

$\sum_0$ 展开于一切无序指数偶$(i, j) \in N, e_i \in \mathscr{P}_\beta, e_j \in \mathscr{P}_\alpha$ 之上而 $x_{ij} = x_{ji}$。

$\sum_1$ 展开于使$(i, \alpha) \in N, e_i \in \mathscr{H}_\beta$ 的一切指数 i 上。

$\sum_2$ 展开于使$(i, \beta) \in N, e_i \in \mathscr{H}_\alpha$ 的一切指数 i 上。

附注　方程组(I)将称为 G 对最大树 T(以及某一 T-浸入)而言的基本方程组。诸 $x_{ij} = x_{ji}((i, j) \in N, e_i, e_j \in T^1)$与 $x_{i\alpha}((i, \alpha) \in N, e_i \in T^1, e_\alpha \in G^1 - T^1)$将称为基本变数。

证　依嵌入定理,$G \subset R^2$ 的一个必要条件为 $\varphi \sim 0$,这里

$$\varphi = \sum \varphi_{\alpha\beta} e_\alpha * e_\beta$$

为相应于某一 T - 浸入 $f: G \propto R^2, f: T \subset R^2$ 的示嵌链。由引理 2,这个必要条件成为:对$(i, j), (i, \alpha) \in D \cup N, e_i, e_j \in T^1$, $e_\alpha \in G^1 - T^1$,有模 2 数 $x_{ij} = x_{ji}$与 $y_{i\alpha}$存在,使

$$\varphi = \sum x_{ij}(e_i + E_i) * (e_j + E_j) + \sum y_{i\alpha}(e_i + E_i) * e_\alpha。 \tag{2}$$

比较两边 $e_i{}^* e_j, e_i{}^* e_\alpha$,与 $e_\alpha{}^* e_\beta(e_i, e_j \in T^1, e_\alpha, e_\beta \in G^1 - T^1$, $(i, j), (i, \alpha)(\alpha, \beta) \in D)$的系数,得

$$x_{ij} = 0, (i, j) \in D, \tag*{①}$$

$$y_{i\alpha}=\sum_{e\alpha\in\mathscr{P}_j}x_{ij}=\sum_{ej\in\mathscr{P}_\alpha}x_{ij},(i,\alpha)\in D, \quad ②$$

$$\varphi_{\alpha\beta}=\sum_{ej\in\mathscr{P}_\beta}y_{i\alpha}+\sum_{ej\in\mathscr{P}_\alpha}y_{i\beta}+\sum_{\substack{ej\in\mathscr{P}_\alpha\\ ej\in\mathscr{P}_\beta}}x_{ij},(\alpha,\beta)\in D。 \quad ③$$

今对$(i,\alpha)\in N$, $e_i\in T^1$, $e_\alpha\in G^1-T^1$引入$x_{i\alpha}$,使

$$y_{i\alpha}=\sum_{ej\in\mathscr{P}_\alpha}x_{ij}+x_{i\alpha},(i,\alpha)\in N。 \quad ④$$

于是对$(\alpha,\beta)\in D$, e_α, $e_\beta\in G^1-T^1$应用①,②,③,即得

$$\begin{aligned}\varphi_{\alpha\beta}=&\sum_{\substack{ei\in\mathscr{P}_\beta\\(ij)\in N}}\sum_{ei\in\mathscr{P}_\alpha}x_{ij}+\sum_{\substack{ei\in\mathscr{P}_\beta\\(i\alpha)\in N}}x_{ij}\\&+\sum_{ei\in\mathscr{P}_\alpha}\sum_{\substack{ei\in\mathscr{P}_\beta\\(ij)\in N}}x_{ij}+\sum_{\substack{ei\in\mathscr{P}_\alpha\\(i\beta)\in N}}x_{j\beta}\\&+\sum_{\substack{ei\in\mathscr{P}_\alpha\\ej\in\mathscr{P}_\beta\\(ij)\in N}}x_{ij}\end{aligned}$$

或

$$\varphi_{\alpha\beta}=\sum_{\substack{ei\in\mathscr{P}_\beta\\ej\in\mathscr{P}_\alpha\\(ij)\in N}}x_{ij}+\sum_{\substack{ei\in\mathscr{P}_\beta\\(i\alpha)\in N}}x_{i\alpha}+\sum_{\substack{ei\in\mathscr{P}_\alpha\\(i\beta)\in N}}x_{i\beta},(\alpha,\beta)\in D。 \quad ⑤$$

显然方程组⑤有解时,方程组①—④亦即(2)有解。故$G\subset R^2$的一个必要条件为方程组⑤有解。但⑤式亦即(Ⅰ)式,故定理得证。

四、基本方程组的分析与简化

方程组(Ⅰ)的分析

设e_α, $e_\beta\in G^1-T^1$, $(\alpha,\beta)\in D$。(Ⅰ)中与(e_α,e_β)相应的方程为

$$\sum_{\substack{ei\in\beta\\ej\in f\alpha\\(ij)\in N}}x_{ij}+\sum_{\substack{ei\in\beta\\(i\alpha)\in N}}x_{i\alpha}+\sum_{\substack{ej\in f\alpha\\(j\beta)\in N}}x_{j\beta}=\varphi_{\alpha\beta},$$

或简写为

$$\textstyle\sum_0+\sum_1+\sum_2=\varphi_{\alpha\beta}。 \quad (\mathrm{I})$$

就$\mathscr{P}_\alpha$与$\mathscr{P}_\beta$的相对关系,可分三种情形。

情形 1 $\mathscr{P}_\alpha, \mathscr{P}_\beta$ 不相遇。

此时$\sum_0, \sum_1, \sum_2$ 都$=0$。由于此时 $\varphi_{\alpha\beta}$必为 0,故相应的方程必为 $0=0$,方程是不足道的。

情形 2 $\mathscr{P}_\alpha, \mathscr{P}_\beta$ 有一顶点 v 公共。

此时设 $\mathscr{P}_\alpha \cup \{e_\alpha\}$在 v 的两个节段为e_a, e_b(规定在 e_α 以 v 为一端点时,e_a, e_b 中有一例如e_b 即作为e_α)。同样设 $\mathscr{P}_\beta \cup \{e_\beta\}$ 在 v 的两个节段为e_r, e_s(规定在 e_β 以 v 为一端点时,e_r, e_s 中有一例如 e_s 即作为e_β)。

若 e_α, e_β 都不以 v 为端点(见图 4),则

$$\sum_0 = x_{ar} + x_{as} + x_{br} + x_{bs},$$

$$\sum_1 = \sum_2 = 0。$$

故相应的方程为

$$x_{ar} + x_{as} + x_{br} + x_{bs} = \varphi_{\alpha\beta}。 \qquad (\text{I})'$$

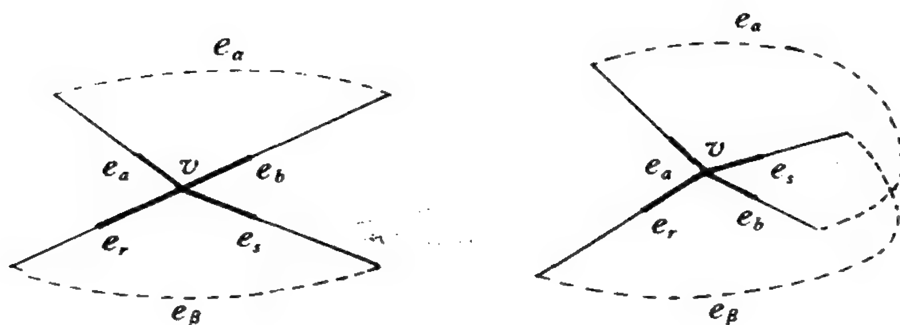


图 4

若 e_β 以 v 为端点(因而 e_α 不再以 v 为端点),则(见图 5)

$$\sum_0 = x_{ar} + x_{br}, \sum_1 = 0,$$

$$\sum_2 = x_{a\beta} + x_{b\beta} = x_{as} + x_{bs}。$$

故相应的方程仍如上述形式。

在 e_α 以 v 为端点时亦然。

情形 3 $\mathscr{P}_\alpha, \mathscr{P}_\beta$ 有一以顶点 v, w 为两端的道路 $\mathscr{P} \subset T$ 公共。

设 $\mathscr{P}$ 在 v, w 处的节段为 e_c, e_t(e_c 也可与 e_t 相同), $\mathscr{P}_\alpha \cup \{e_\alpha\}$ 在 v 处的节段除 e_c 外设为 e_a(规定在 e_α 以 v 为端点时 e_a 即作为 e_α), 在 w 处的节段除 e_t 外设为 e_r(规定在 e_α 以 w 为端点时 e_r 即作为 e_α)。同样, $\mathscr{P}_\beta \cup \{e_\beta\}$ 在 v, w 处的节段除 e_c, e_t 外设为 e_b, e_s(规定在 e_β 以 v 或 w 为端点时 e_b 或 e_s 亦即为 e_β)。

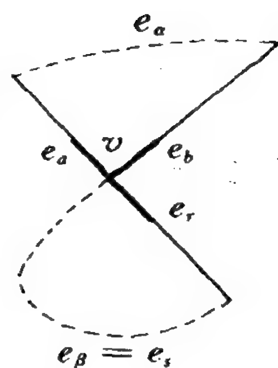


图 5

若 e_α, e_β 都不以 v, w 为端点(见图 6), 则

$$\sum_0 = x_{rs} + x_{rt} + x_{st} + x_{ab} + x_{ac} + x_{bc},$$

$$\sum_1 = \sum_2 = 0。$$

故相应的方程为

$$x_{rs} + x_{rt} + x_{st} + x_{ab} + x_{ac} + x_{bc} = \varphi_{\alpha\beta}。 \qquad (\text{I})''$$

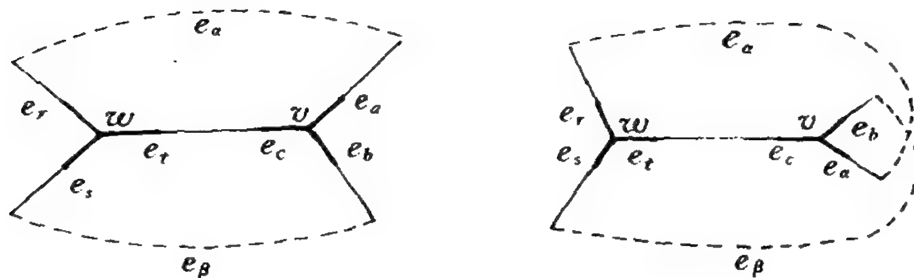


图 6

其次, 若 e_β 以 v, w 为端点(因而 e_α 不能以 v, w 为端点), 则(见图 7)

$$\sum_0 = x_{rt} + x_{ac}, \sum_1 = 0,$$

$$\begin{aligned}\sum_2 &= x_{r\beta} + x_{t\beta} + x_{a\beta} + x_{c\beta} \\ &= x_{rs} + x_{st} + x_{ab} + x_{bc}。\end{aligned}$$

因而相应方程仍如上式(I)″。

同样，若 e_α 以 v 为端点，e_β 以 w 为端点(见图 8)，则

$$\sum_0 = x_{rt} + x_{bc},$$

$$\sum_1 = x_{ba} + x_{ca} = x_{ab} + x_{ac},$$

$$\sum_2 = x_{r\beta} + x_{t\beta} = x_{rs} + x_{st},$$

方程仍如(I)″。

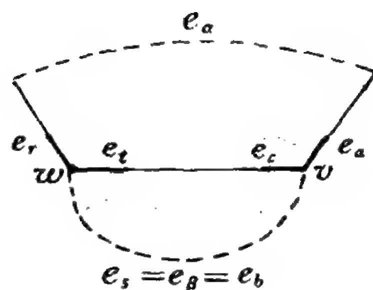


图 7

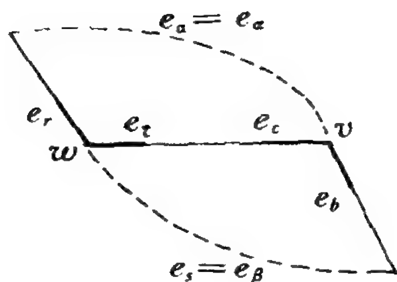


图 8

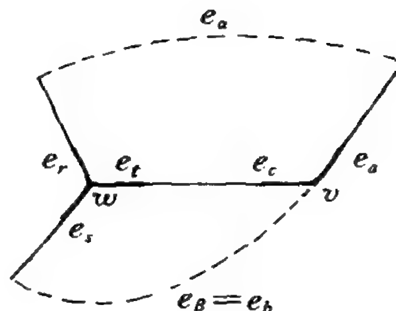


图 9

最后，若 e_β 以 v 为端点，而 e_α 不以 v, w 为端点(见图 9)，则

$$\sum_0 = x_{rs} + x_{rt} + x_{st} + x_{ac},$$

$$\sum_1 = 0,$$

$$\sum_2 = x_{a\beta} + x_{c\beta} = x_{ab} + x_{bc},$$

方程仍如(I)″。

其余情形可从以上情形互易 e_α, e_β 或 v, w 而得出。所得方程恒如(I)″。

总结起来，得

命题 方程组(I)的每一方程，除不足道者(作 0=0 形式)外，必如形式

$$x_{ar} + x_{as} + x_{br} + x_{bs} = \varphi_{\alpha\beta} \qquad (\text{I})'$$

或

$$x_{rs}+x_{rt}+x_{st}+x_{ab}+x_{ac}+x_{bc}=\varphi_{\alpha\beta}。\qquad (\text{I})''$$

为了显溪起见，特将各种情形列表如下（两种情形可从对易例如将 e_α, e_β 或 v, w 对易而得出者不同时并列）：

情形		$\sum_0$	$\sum_1$	$\sum_2$	方程形式
I	e_α e_β	0	0	0	0=0
II	e_α e_a e_b e_s e_β	$x_{ar}+x_{as}+x_{br}+x_{bs}$	0	0	(I)′
	e_α e_a e_b e_r $e_\beta=e_s$	$x_{ar}+x_{br}$	0	$x_{a\beta}+x_{b\beta}=x_{as}+x_{bs}$	

（续）

情形		Σ_0	Σ_1	Σ_2	方程形式
	e_α, e_r, e_a, e_t, e_c, e_b, e_s, e_β	$x_{rs}+x_{rt}+x_{st}+x_{ab}+x_{ac}+x_{bc}$	0	0	(Ⅰ)″
	e_α, e_r, e_t, e_c, e_a, $e_s=e_\beta=e_b$	$x_{rt}+x_{ac}$	0	$x_{rB}+x_{t\beta}+x_{a\beta}+x_{CB}=x_{rs}+x_{st}+x_{ab}+x_{bc}$	
Ⅲ	$e_\alpha=e_a$, e_r, e_b, e_t, e_c, $e_s\Rightarrow e_\beta$	$x_{rt}+x_{bc}$	$x_{ba}+x_{ca}=x_{ab}+x_{ac}$	$x_{r\beta}+x_{t\beta}=x_{rs}+x_{st}$	(Ⅰ)″
	e_α, e_r, e_t, e_c, e_a, e_s, $e_\beta=e_b$	$x_{rs}+x_{rt}+x_{st}+x_{ac}$	0	$x_{a\beta}+x_{c\beta}=x_{ab}+x_{bc}$	

树根与多余变数

为了将平面图具体作出，须要进一步将方程简化。为此引入一些概念如下。

在最大树 T 中任取一个节梢 0，称之为树根。对任一顶点 $v \in G^0 = T^0$，在 T 中有一唯一的通道连接 v 与 O，称为到 v 的干道，记之为 $\mathscr{P}_v$. 设 e_i 是 $\mathscr{P}_v$ 上在 v 处的节段，则对任一在 v 处的棱 $e_i \in G^1$，x_{ij} 将称为一多余变数。对在 v 处而不同于 e_i 的棱 $e_j, e_k \in G^1$，当 e_j, e_k 都 $\in G^1 - T^1$ 时，x_{jk} 称补充变数，否则 x_{jk} 称主要变数。此外，对一 T - 浸入 f，除相应示嵌链

$$\varphi = \sum \varphi_{\alpha\beta} e_\alpha{}^* e_\beta$$

中出现的一组系数 $\varphi_{\alpha\beta}((\alpha, \beta) \in D, e_\alpha, e_\beta \in G^1 - T^1)$ 者外，对有一个顶点公共的 $e_\alpha, e_\beta \in G^1 - T^1$，我们也将定义 $\varphi_{\alpha\beta}$ 为 $f(e_\alpha)$，$f(e_\beta)$ 除去在公共端点外的交点数。

扩充方程组

原来基本方程组(Ⅰ)(或(Ⅰ)′，(Ⅰ)″)中出现的变数 x_{ij}，都是主要变数或多余变数。今设 $e_\alpha, e_\beta \in G^1 - T^1$ 而 e_α, e_β 有一公共顶点 v。此时 $\mathscr{P}_\alpha$ 与 $\mathscr{P}_\beta$ 可只相遇于 v 亦可有一两端为 v, w 的通路 $\mathscr{P}$ 公共。在第一情形(见图 10)令 $\mathscr{P}_\alpha$ 中在 v 处的节段为 e_b，$\mathscr{P}_\beta$ 中在 v 处的节段为 e_s，又置 $e_a = e_\alpha, e_r = e_\beta$，此时我们将引入方程

(Ⅰ)′$_0$ $\qquad x_{ar} + x_{as} + x_{br} + x_{bs} = \varphi_{\alpha\beta}$。

其次在第二情形(见图 11)，我们将设 $\mathscr{P}$ 上在 v, w 处的节段各为 e_c, e_t，在 $\mathscr{P}_\alpha$ 上 w 处不同于 e_t 的节段为 e_r，而在 $\mathscr{P}_\beta$ 上 w 处不同于 e_t 的节段为 e_s。在 e_α 或 e_β 以 w 为端点时，则置 $e_r = e_\alpha$ 或 $e_s = e_\beta$。又置 $e_a = e_\alpha, e_b = e_\beta$。此时将引入方程

(Ⅰ)″$_0$ $\qquad x_{rs} + x_{rt} + x_{st} + x_{ab} + x_{ac} + x_{bc} = \varphi_{\alpha\beta}$。

由于这些方程与原来的基本方程(Ⅰ)′，(Ⅰ)″形式相同，且每个方程中恰含有一个补充变数(即 x_{ar} 与 x_{ab})，而不同的方程

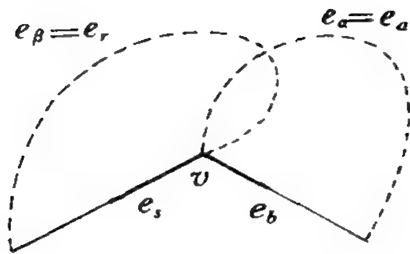


图 10

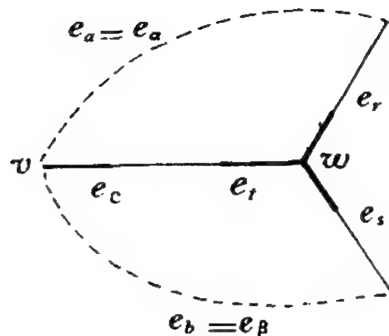


图 11

所含的补充变数也不同，因之，如果将这些方程加入原来的基本方程组(Ⅰ)，方程有解与否的性质将并不因之而变。

我们将称添加$(\text{Ⅰ})_0'$，$(\text{Ⅰ})_0''$后的方程组为扩充方程组，并记之为$(\tilde{I})$，$(\tilde{I})'$，$(\tilde{I})''$等。

引理 对最大树 T 与根 O 而言，对每一节点处 $e_i, ej \in G^1$ 引入变数 x_{ij}' 并置

$$\left.\begin{aligned}[i,j,k] &= x_{ij} + x_{ik} + x_{jk},\\ [i,j,k]' &= x_{ij}' + x_{ik}' + x_{jk}'\end{aligned}\right\} e_i, e_j, e_k \text{ 在同一节点处时。}$$

则当 x_{ij} 为已知数时，方程组

$$\begin{cases}[i,j,k]' = [i,j,k],\\ x_{rs}' = 0, (x_{rs}\text{为多余变数时})\end{cases}$$

有解。

证 在 x_{rs} 为主要变数或补充变数而 e_r, e_s 的公共节点非树根 O 时，令自 e_r, e_s 的公共节点到 O 的通道在此节点处的节段为 e_t，而置

$$x_{rs}' = [r,s,t],$$

在 x_{rs} 为主要变数或补充变数而 e_r, e_s 的公共节点即树根 O 时，置 $x'_{rs} = x_{rs}$。在 x_{rs} 为多余变数时，则置

$$x_{rs}' = 0。$$

试证这样的 x_{rs}' 满足上述方程组。为此，设 $e_i, e_j, e_k \in G^1$ 的公

共节点与 O 不同而自此至 O 的通道在此节点处的节段为 e_l。若 e_l 即为 e_i, e_j, e_k 之一例如 $e_l = e_k$，则将有

$$[i,j,k]' = [i,j,l]' = x'_{il} + x'_{jl} + x'_{ij} = x'_{ij} = [i,j,l]。$$

若 e_l 不与 e_i, e_j, e_k 相同，则将有

$$\begin{aligned}[i,j,k]' &= [i,j,l]' + [i,k,l]' + [j,k,l]' \\ &= [i,j,l] + [i,k,l] + [j,k,l] \\ &= [i,j,k]。\end{aligned}$$

若 $e_i, e_j, e_k \in G^1$ 以 O 为公共节点，则 $[i,j,k]' = [i,j,k]$ 甚显然，故引理得证。

基本定理Ⅱ 对 G 的最大树 T 与选定的根 O 而言，若基本方程组（Ⅰ）可解，则必有一解答，它的多余变数都 $=0$。对于扩充方程组（$\bar{I}$）亦然。

证 已知方程组（Ⅰ）等价于方程组（Ⅰ）′与（Ⅰ）″。但（Ⅰ）′与（Ⅰ）″各可写作如下形式：

$$[ars] + [brs] = \varphi_{\alpha\beta}, \qquad (\text{Ⅰ})'$$

$$[rst] + [abc] = \varphi_{\alpha\beta}。 \qquad (\text{Ⅰ})''$$

由引理，上方程组有解时，下方程组也有解：

$$\begin{cases} [ars]' + [brs]' = \varphi_{\alpha\beta}, \\ [rst]' + [abc]' = \varphi_{\alpha\beta}, \\ x'_{rs} = 0(x_{rs} = \text{多余变数时}), \end{cases}$$

亦即

$$\begin{cases} x'_{ar} + x'_{as} + x'_{br} + x'_{bs} = \varphi_{\alpha\beta}, \\ x'_{rs} + x'_{rt} + x'_{st} + x'_{ab} + x'_{ac} + x'_{bl} = \varphi_{\alpha\beta}, \\ x'_{rs} = 0(x_{rs} = \text{多余变数时}), \end{cases}$$

或

$$\begin{cases} \sum_0' + \sum_1' + \sum_2' = \varphi_{\alpha\beta}, \\ x'_{rs} = 0(x_{rs} = \text{多余变数时}) \end{cases}$$

有解，其中$\sum'$即在$\sum^0$中将x_{ij}都易为x'_{ij}而得。因之当(Ⅰ)有解时，(Ⅰ)必有一解答，它的多余变数都$=0$。对$(\tilde{I})$亦然，证毕。

定义 基本方程组的一组解答将称为T-侵入f的一个基本解(对T与O而言)。同样，扩充方程组的任意一组解答将称为f的一个扩充解，多余变数都$=O$的解答则将称为f的简化解。

附注：如果将线性图G改变为G'，使对每一$e_a \in G^1 - T^1$，在e_a上添入两个新顶点而分之为三段$e_{a_1}, e_{a_2}, e_{a_3}$，其中$e_{a_3}$为$e_a$上中间的一段，将$e_{a_1}, e_{a_2}$添入$T$而成$T'$，则$T'$是$G'$的一个最大树，于是对$G', T'$而言，方程组(Ⅰ)与$(\tilde{I})$除不足道者外将只出现图1与图3两种情形，而不再出现图2与图4—8等情形。若对G'的每一变数，例如$x_{a_{1i}}$的下角出现a_1, a_2, a_3者都易a_λ为a，即可重行获得对G, T的相应方程组，因之前面分析的许多情形实际上可以统一起来。

五、平面图的具体作法

我们已在四中给出了$G \subset R^2$的必要条件：基本方程组(Ⅰ)或扩充方程组$(\tilde{I})$有解。本节将证明，这个条件也是充分的，并将给出从方程组(Ⅰ)或$(\tilde{I})$的一个解答具体作图的方法。

调整

为了要将一具有平面性的线性图G在平面中具体作出，须将一不完全是嵌入的T-浸入逐步改变成一嵌入，为此引入调整概念如下。

设$f: G \propto R^2$是一T-浸入，$v \in G^0 = T^0$的干道$\mathscr{P}_v$中在

v 处的节段为e_s*。设 e_i，$e_j \in G^1$ 为在 v 处不同于 e_s 的任意两棱。设在 f 之下，自 e_s 起，绕 v 作顺钟向旋转时，在 v 处的诸棱依次为

$$e_s,\ e_{p1},\ \cdots,\ e_{p\lambda},\ e_i,\ e_{q1},\ \cdots,\ e_{q\mu},\ e_j,\ e_{r1},\ \cdots,\ e_{r\gamma}。$$

今另作一 T－浸入 $g: G \subset R^2$，满足以下诸条件：

（1）在 g 之下，自 e_s 起，绕 v 作顺钟向旋转时，在 v 处的诸棱依次为

$$e_s,\ e_{p1},\ \cdots,\ e_{p\lambda},\ e_j,\ e_{q1},\ \cdots,\ e_{q\mu},\ e_i,\ e_{\gamma 1},\ \cdots,\ e_{rv}。$$

（2）在任一不同于 v 的顶点 w 处，绕 w 作顺钟向旋转时，在 w 处的诸棱在 f 下与在 g 下的次序相同。

这样的 g 必然存在，我们将称之为将 e_i，e_j 互易的 f 的一个调整。

调整定理（特殊形式）　设 $f: G \subset R^2$ 是一 T－浸入。设 e_i，$e_j \in G^1$ 都在节点 v_0 处，不位于干道 $\mathscr{P}_{v^0}$上，且 e_i，e_j 在 f 下处于相邻位置。今调整 f 为 T－浸入 g，使 e_i，e_j 互易。则当 f 有一组扩充解 $x_{\lambda\mu} = c_{\lambda\mu}$时，$g$ 也有一组扩充解 $x_{\lambda\mu} = d_{\lambda\mu}$，这里

$$d_{\lambda\mu} = \begin{cases} c_{\lambda\mu} + 1, (\lambda, \mu) = (i, j), \\ c_{\lambda\mu} (\lambda, \mu) \neq (i, j)。\end{cases}$$

再者，若$(c_{\lambda\mu})$对根 O 而言是f 的一组简化解，则$(d_{\lambda\mu})$也是 g 的一组简化解。

证　设 f, g 的示嵌链各为 $\varphi = \Sigma \varphi_{\alpha\beta} e_\alpha{}^* e_\beta$，$\psi = \Sigma \psi_{\alpha\beta} e_\alpha{}^* e_\beta$，则$(c_{\lambda\lambda})$为下扩充方程组的解：

$$\begin{cases} (\tilde{I})'f \quad x_{ar} + x_{as} + x_{br} + x_{bs} = \varphi_{\alpha\beta}, \\ (\tilde{I})''f \quad x_{rs} + x_{rt} + x_{st} + x_{ab} + x_{ac} + x_{bc} = \varphi_{\alpha\beta}。\end{cases}$$

---

* 在 v = 树根 O 时，可在 G 中添入一棱e_0 以 O 为其一端，于是 e_s 将取为e_0。

我们须证$(d_{\lambda\mu})$为下扩充方程组的解：

$$\begin{cases}(\widetilde{I})'_g & x_{ar}+x_{as}+x_{br}+x_{bs}=\psi_{\alpha\beta},\\ (\widetilde{I})''_g & x_{rs}+x_{rt}+x_{st}+x_{ab}+x_{ac}+x_{bc}=\psi_{\alpha\beta}。\end{cases}$$

在情形$(\widetilde{I})'$中，若$v\neq v_0$或$v=v_0$但(a, r)，(a, s)，(b, r) (b, s)都$\neq(i, j)$，则(e_a, e_b, e_r, e_s)绕v的顺序在f下与在g下二者相同，或至多e_a，e_b互易，或e_r，e_s互易，不论何时都有$\psi_{\alpha\beta}=\varphi_{\alpha\beta}$。另一面又有$d_{ar}=c_{ar}$，…，$d_{bs}=c_{bs}$。因之当$(c_{\lambda\mu})$满足$(\widetilde{I})'f$时，$(d_{\lambda\mu})$将满足相应的$(\widetilde{I})'g$。反之，若$v=v_0$，而在$(a, r)$，$(a, s)$，$(b, r)$，$(b, s)$中有一，例如$(a, r)=(i, j)$，因而在$f$下绕$v$的诸棱中$e_a$与$e_r$相邻（见图12），则将有$\psi_{\alpha\beta}=\varphi_{\alpha\beta}+1$，$d_{ar}=c_{ar}+1$，$d_{as}=c_{as}$，$d_{br}=c_{br}$，$d_{bs}=c_{bs}$。因而当$(c_{\lambda\mu})$满足$(\widetilde{I})'_f$时，$(d_{\lambda\mu})$，仍将满足相应的$(\widetilde{I})_g{}'$。

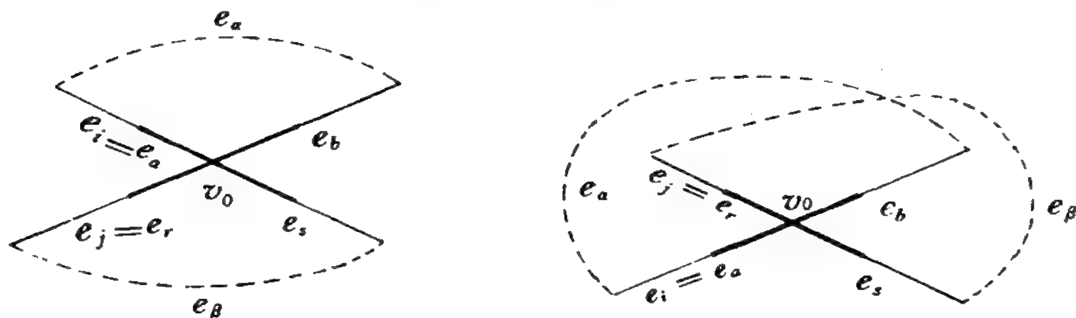


图 12

其次，在情形$(\widetilde{I})''$中，若v，w都$\neq v_0$或$v=v_0$但(a, b)，(a, c)，(b, c)都$\neq(i, j)$，或$w=v_0$但(r, s)，(r, t)，(s, t)都$\neq(i, j)$，则显有$\psi_{\alpha\beta}=\varphi_{\alpha\beta}$，$d_{rs}=c_{rs}$，…。故当$(c_{\lambda\mu})$满足$(\widetilde{I})_f''$时，$(d_{\lambda\mu})$将满足相应的

$(\tilde{I})_g''$。反之，若 v，w 中有一，例如 $v=v_0$，又在 $(a,\ b)$，$(a,\ c)$，$(b,\ c)$ 中有一，例如 $(a,\ c)=(i,\ j)$，则将有（见图 13）：

$$\psi_{\alpha\beta}=\varphi_{\alpha\beta}+1,\quad d_{ac}=c_{ac}+1,\quad d_{ab}=c_{ab},\quad \cdots。$$

故 $(c_{\lambda\mu})$ 满足 $(\tilde{I})''_f$ 时，$(d_{\lambda\mu})$ 仍满足相应的 $(\tilde{I})_g''$。其余情形亦同。这证明了定理的前一部分。

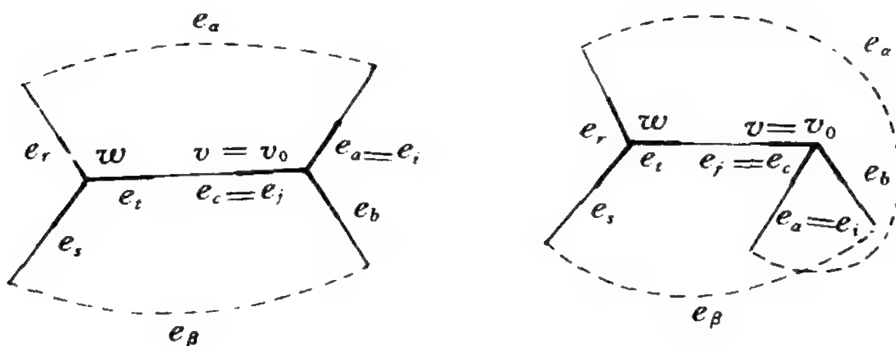


图 13

最后，因 e_i，e_j 都不在干道 $\mathscr{P}_{v0}$ 上，故在简化解当 $x_{\lambda\mu}$ 为多余变数而 $c_{\lambda\mu}=0$ 时，同样亦有 $d_{\lambda\mu}=0$，因之 $(d_{\lambda\mu})$ 亦为一简化解。证毕。

调整定理（一般形式） 设 $f:G\subset R^2$ 是一 T－浸入，设 e_i，$e_j\in G^1$ 在节点 v_0 处，不位于干道 $\mathscr{P}_{v0}$ 上，且在 f 下 e_i，e_j 之间不含通道 $\mathscr{F}_{v0}$ 的部分恰有 μ 个棱 $e_{q1},\cdots,e_{q\mu}$ 以 v_0 为端点。今调整 f 为 T－浸入 g，使 e_i，e_j 互易。则当 f 有一组扩充解 $x_{\lambda\mu}=c_{\lambda\mu}$ 时，g 也有一组扩充解 $x_{\lambda\mu}=d_{\lambda\mu}$，这里

$$d_{\lambda\mu}=\begin{cases}c_{\lambda\mu}+1,(\lambda\mu)=(i,j),(i,q_1),\cdots,\\(i,q_\mu),(j,q_1),\cdots\text{或}(j,q_\mu)。\\c_{\lambda\mu},\quad\text{在其他情形。}\end{cases}$$

再者，若 $(c_{\lambda\mu})$ 对根 O 而言是 f 的一组简化解，则 $(d_{\lambda\mu})$ 也是 g 的一组简化解。

证　迭次应用特殊形式的调整定理即得。

合格化

设 $e_a \in G^1 - T^1$, $C_a = \{e_a\} \cup \mathscr{F}_a$, $f: G \subset R^2$ 为 T－浸入。假设对应于 T 以及根 O 而言，f 的扩充方程组$(\widetilde{I})$可解且有一组简化解$(c_{\lambda\mu})$。如果下面的条件满足，则称 f 对 e_a 合格而$(c_{\lambda\mu})$是 f 对 e_a 的一组合格解。条件是：

(C)　　对任一 $e_i \in (C_a)^1$，以及任一与 e_i 有公共顶点的 $e_j \in G^1$，有 $c_{ij} = 0$。

如果 f 对 e_a 不合格，因而有 $e_i \in (C_a)^1$，以及与 e_i 有公共顶点的 $e_j \in G^1$ 使 $c_{ij} = 1$。依据调整定理，如果将每一对这样的 e_i，e_j 互易或改变 $(c_{\lambda\mu})$，则可调整 f 为一 T－浸入 g，使 g 对 e_a 合格，且从 $(c_{\lambda\mu})$ 依调整定理所得对 g 的一组解答 $(d_{\lambda\mu})$ 是 g 对 e_a 的一组合格解。这个从 f 调整为 g 的手续将称为 f 对 e_a 的合格化。显然，在调整过程中可取这样的 g，使 g 与 f 在 C_a 上相合。

引理　设 $e_a \in G^1 - T^1$，$f: G \subset R^2$ 是－T－浸入，对 e_a 合格。则可改变 f 为一 T－浸入 g，使 $g \equiv f / T \cup \{e_a\}$，且

(1) 对任一 $\neq e_a$ 的 $e_\beta \in G^1 - T^1$，e_β 与 e_a 在 g 下至多在公共顶点处相遇。

(2) 对任一 $\neq e_a$ 的 $e_\beta \in G^1 - T^1$，g 与 f 在 e_β 的两端的邻近相合。

证　记 e_β 的两端为 a_β，b_β. 若 a_β（或 b_β）不在 $\mathscr{P}_a$ 上，则置 $a'_\beta = a_\beta$（或 $b'_\beta = b_\beta$）。若 a_β（或 b_β）在 $\mathscr{P}_a$ 上，则取 e_β 内部充分邻近 a_β（或 b_β）的一点为 a'_β（或 b'_β）。

今设 $\mathscr{P}_\beta$ 不与 $\mathscr{P}_a$ 相遇，则 $a'_\beta = a_\beta$ 与 $b'_\beta = b_\beta$ 在 f 下必同在 C_a 之内或同在 C_a 之外。

若 $\mathscr{P}_\beta$ 与 $\mathscr{P}_a$ 相遇于一点，则对 e_α，e_β 的扩充方程有形式

$(\tilde{I})'$。由于 f 对 e_α 合格，故方程左端的诸 $x_{ij}=c_{ij}$ 都 $=0$，因而 $\varphi_{\alpha\beta}=0$，故此时在 f 下 $a'\beta$ 与 b'_β（只须充分接近 a_β，b_β）仍同在 c_α 之内或同在其外。

若 $\mathscr{P}_\beta$ 与 $\mathscr{P}_\alpha$ 相遇于一段通道，则相应的扩充方程有形式 $(\tilde{I})''$。此时由于 f 对 e_α 合格，方程左端的诸 $x_{ij}=c_{ij}$ 都 $=0$，而仍得同样结论。

故不论何时，在 f 下 a'_β 与 b'_β 都将同在 c_α 之内或外。因而可改变 f 在 $a'_\beta b'_\beta$ 的部分，使在所得 g 之下 $a'_\beta b'_\beta$ 不再与 e_α 相遇，如所欲证。

引理 设 $e_\alpha \in G^1-T^1$，以及 $T-$浸入 f，$g: G \subset R^2$ 如上一引理，这里 f、g 都对 e_α 合格。于是 g 对 e_α 有一组合格（简化）解（$c_{\lambda\mu}$）不仅满足（C），且满足下述条件

$(C)_\alpha{}^*$ 对任意在 $(C_\alpha)^1$ 上有公共顶点的 e_i，$e_j \in G^1$，只须 e_i，e_j 中一在 C_α 之内而一在 C_α 之外（除公共顶点外），即有 $C_{ij}=0$。

证 已知 g 有一组合格（简化）解（$c'_{\lambda\mu}$）满足条件（C）。今依下方式改变（$c'_{\lambda\mu}$）为（$c_{\lambda\mu}$）：若与 c_{ij} 对应的 e_i，e_j 如 $(C)_\alpha{}^*$ 中所示，则令 $c_{ij}=0$，否则置 $c_{ij}=c'_{ij}$。于是（$c_{\lambda\mu}$）满足条件（C）与 $(C)_\alpha{}^*$，我们只须证明（$c_{\lambda\mu}$）满足 g 的扩充方程组即可。

设方程中不出现任何 x_{ij}，此外 e_i，e_j 如 $(C)_\alpha{}^*$ 中所示时，对方程中出现的诸 $x_{\lambda\mu}$ 将有 $c_{\lambda\mu}-c'_{\lambda\mu}$。由于（$c'_{\lambda\mu}$）满足这一方程，故（$c_{\lambda\mu}$）亦然。

其次试考虑任一对以 C_α 上一点 v 为公共顶点的 e_i，$e_j \in G^1$，此外 e_i 在 c_α 之内而 e_j 在 c_α 之外（除公共顶点 v 以外）。于是在 g 的扩充方程组中有 x_{ij} 的方程必如下产生：有 e_β，$e_r \in G^1-T^1$ 使 $\mathscr{P}_\beta$ 含有 e_i 而 $\mathscr{P}_r$ 含有 e_j。由上一引理的证明可知

在 g 下必有 e_β 在 C_α 之内而 c_r 在 C_α 之外（除可能在 $\mathscr{P}_\alpha$ 上的端点以外），且 $\varphi_{\beta r}=0$。

先设 $\mathscr{P}_\beta$，$\mathscr{P}_r$ 有一通道 $\mathscr{P}_{\beta r}$ 公共，则必有 $\mathscr{P}_{\beta r}\subset\mathscr{P}_\alpha$ 且 $\mathscr{P}_{\beta r}$ 一端为 v，而另一端设为 w。相应的方程必如形式

$$x_{ij}+x_{ik}+x_{jk}+x_{rs}+x_{rt}+x_{st}=\varphi_{\beta r}=0。\qquad (1)$$

此外 e_k，e_t 位于通道 $\mathscr{P}_{\beta r}$ 上，e_r 在 $\mathscr{P}_\beta\cup\{e_\beta\}$ 上，e_s 在 $\mathscr{P}_r\cup\{e_r\}$ 上。e_k 以 v 为一端而 e_r，e_s，e_t 以 w 为一端。仍由上引理，在 g 下 e_r 在 C_α 之内或 $\mathscr{P}_\alpha$ 之上。而 e_s 在 C_α 之外或 $\mathscr{P}_\alpha$ 之上(端点可能除外)。由于($c_{\lambda\mu}$)满足$(C)_\alpha^*$，故不论何时都有

$$c_{ij}=c_{ik}=c_{jk}=c_{rs}=C_{rt}=c_{st}=0。$$

因而（$c_{\lambda\mu}$）满足（1）。

其次，设 $\mathscr{P}_\beta$，$\mathscr{P}_r$ 只有 v 点公共，则相应的方程形如

$$x_{ij}+x_{il}+x_{kj}+x_{kl}=\varphi_{\beta r}=0。\qquad (2)$$

此外 e_k，e_l 各在 $\mathscr{P}_\beta\cup\{e_i\}$ 与 $\mathscr{P}_r\cup\{e_j\}$ 上而以 v 为端点。与前同样不论何时都有 $c_{ij}=c_{il}=c_{kj}=c_{kl}=0$，因而($c_{\lambda\mu}$)满足(2)。

由此知（$c_{\lambda\mu}$）满足 g 的扩充方程组，如所欲证。

作图法

设有线性图 G。今取一最大树 T 与根 O。任取一 T－浸入 $f:G\subset R^2$。假设其相应示嵌链 $\varphi\sim0$。依基本定理Ⅱ，扩充方程组（$\tilde{I}$）有一简化解（c_{ij}）。今将诸 $e_\alpha\in G^1-T^1$ 排成一次序 $e_{\alpha1}$，$e_{\alpha2}$，…，$e_{\alpha s}$。依调整定理从（c_{ij}）出发可改变 f 为一 T－浸入 $f_1:G\subset R^2$，使对 $e_{\alpha1}$ 合格，且由上两引理可使 f_1 满足该二引理所说的条件，于是在 f_1 下 $e_{\alpha2}$，…，$e_{\alpha s}$ 都与 $e_{\alpha1}$ 至多相遇于其公共顶点，且与 f_1 相应的扩充简化解（c_{ij}^1）满足条件（C）$_{\alpha1}^*$。仍依调整定理从（c_{ij}^1）出发可改变 f_1 为一 T－浸入 $f_2:G\subset R^2$，以及一组相应的扩充简化解（C_{ij}^2），使对 $e_{\alpha2}$ 合格，且满足上两引理所说的条件（1）与（C）$_{\alpha_2}^*$。由调

整过程知可取 f_2，使对 e_{a1}仍然合格且（C_{ij}^2）仍满足（C)$_{a_1}^*$。于是在 f_2 下 e_{a3}，…，e_{as}都与e_{a1}，e_{a2}至多相遇于其公共顶点，e_{a1}与 e_{a2}亦然。改变 f_2 使对 e_{a3}合格，且使 f_2 满足（1)，又使相应扩充简化解（C_{ij}^3）满足条件（C)$_{a3}^*$，依次进行最后即得一 T－浸入 $f_s: G \subset R^2$ 对任一 e_{ak}都合格，且任两 e_{ak}与e_{a1}至多只相遇于其公共顶点。这样的一个 T－浸入实为一嵌入。因而 G 已嵌入于平面 R^2 中。

根据二中嵌入定理与上作图法，可得下面完全形式的

嵌入定理 * 线性图 G 具有平面性的充要条件为示嵌类

$$\Phi^2 = 0。$$

同样，三的基本定理Ⅰ亦可重述之为：

基本定理Ⅰ G 有平面性的充要条件为，对最大树 T 而言，基本方程组（Ⅰ）有解。

六、一些估计

在以下，G 的顶点数与棱数将各设为 m 与 n。

将前数节解法在计算机上实施时，最占存储量的是将线性方程组存入计算机的这一部分。如果未知数的个数是 A，则所占用计算机的字码(bits)数估计将为 A^2(方程的个数并不是主要的)。如果依照二所提及的方程组，A 将约略等于 $m \cdot n$，但对于§3—4 中所引入的简化了的基本方程组，此数将大为减少。而且，我们只要估计主要变数的个数即足，因为其余的变数可直接由主要变数表达出来，而在计算机中不需要占用什么存储单元(这些方程直接可从 G 与 T 读出)。

---

* 嵌入定理的原来证明（见（[7]）须依据 Kuratowski 定理（[4]）。但本文的证明并不须要。

为作这一估计，我们将先考虑下述特殊情形：

G 的每一顶点都是三叉点，即每一顶点处都恰有三个棱以此为端点。

今作一最大树 T，取树根 O。设非节梢的节点数为 r，则节梢的个数为 $m-r$，其中恰有 T 的三个节段以之为端点的节点数设为 s，则恰有两个节段以之为端点的节点数将为 $r-s$。由于 T 中节段的个数是 $m-1$，故有

$$m-r+2(r-s)+3s=2(m-1)$$

或

$$r+s=m-2。$$

由于每一非节梢的节点恰恰给出一个主要变数，而在树根处又恰有两个主要变数，因而主要变数的个数 A 将为

$$A=r+2=m-s\leqslant m。$$

由此并知选择 T 时应使 T 的三叉点数 s 尽量的大，以使 A 尽量的小。

在一般情形，我们将易 G 为一只有三叉点的线性图 G'，使 G 中每一个 k 叉点($k\geqslant 3$)易为 G' 中 $k-2$ 个三叉点，图14所示。

设 G 中无二叉点，而 k 叉点的个数为 m_k($k\geqslant 3$)，则 G' 中三叉点的个数将为

$$m'=\sum(k-2)m_k。$$

因 G 的顶点数为 m，棱数为 n，故有

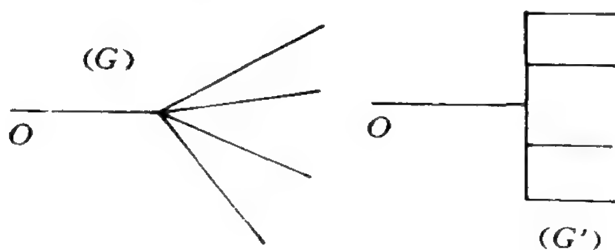


图 14

$$m=\sum m_k,$$

$$2n=\sum km_k。$$

由此得 $$m'=2(n-m)。$$

对于 G'的主要变数个数因之将为

$$A'\leqslant 2(n-m)。$$

由于在一个电路中线路可以假定是没有电阻的，因之与任一电路相应的线性图 G 恒可易为只有三叉点的线性图 G'而无损于一般性。故恒可取上面的公式来估计需解线性方程组的未知数个数，由此以估计计算机中需占用的存储量。

七、杂　　例

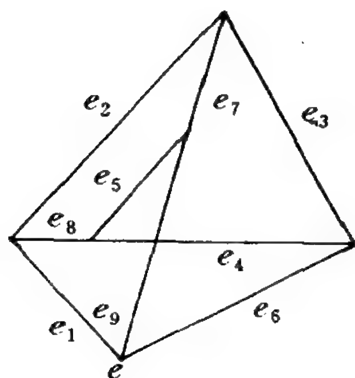
例 1　kuratowski 第一类非平面性线性图(见图 15)。

取最大树 T，根 O 及 T－浸入 f 如下图。主要变数为

$$x_{28},x_{37}x_{46},x_{58}，以及\ O\ 处\ x_{16},x_{19}。$$

基本方程组(已简化)为

$$I_f(e_6,e_7)=1=x_{37}+x_{46},$$
$$I_f(e_6,e_8)=0=x_{28}+x_{46},$$
$$I_f(e_7,e_8)=0=x_{37}+x_{58},$$
$$I_f(e_8,e_9)=0=x_{28}+x_{58},$$



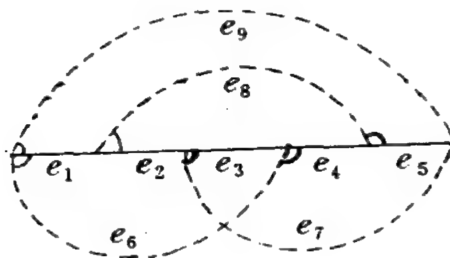


图 15

$$I_f(e_6,e_9)=0=x_{46}+x_{16}+x_{19},$$

以及

$$I_f(e_7,e_9)=0=x_{37}。$$

因方程组矛盾，故原图形是非平面性的。但若除去 e_9，则相应方程组有解(简化解)为

$$(x_{28},x_{37},x_{45},x_{58})=(0,1,0,1),$$

$$或=(1,0,1,0)。$$

而 x_{16}任意。今取 $x_{16}=0$ 而从这两解进行合格化调整，可得两个具体嵌入各如下二图所示：

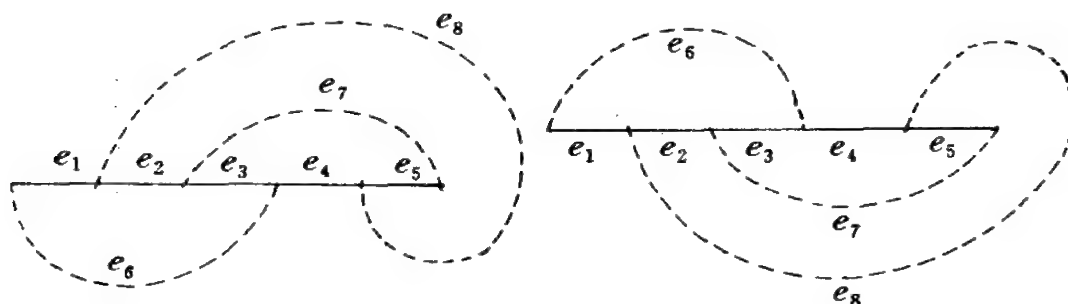


图 16

例 2　Kuratowski 第二类非平面性线性图(见图 17)。

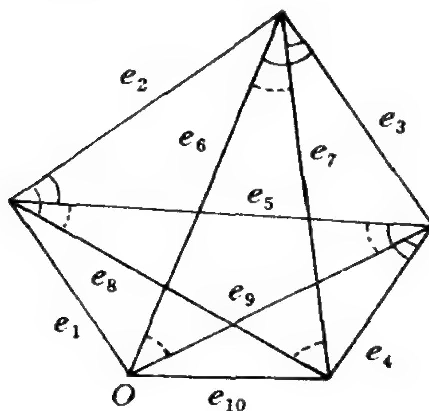


图 17

取最大树 T，树根 O，与 T－浸入 f 如图。已简化的基本方程组为

$$I_f(e_5,e_6)=1=x_{25}+x_{36},$$

$$I_f(e_5, e_7) = 1 = x_{45} + x_{37},$$
$$I_f(e_6, e_8) = 1 = x_{28} + x_{36},$$
$$I_f(e_7, e_9) = 1 = x_{37} + x_{49},$$
$$I_f(e_8, e_9) = 1 = x_{28} + x_{49},$$
$$I_f(e_5, e_{10}) = 0 = x_{25} + x_{45}。$$

由此知 e_{10}须除去以平面化，记所余图形为 G'。相应基本解为

$$(x_{25}, x_{28}, x_{36}, x_{37}, x_{45}, x_{49}) = (0, 0, 1, 0, 1, 1)$$

或

$$= (1, 1, 0, 1, 0, 0),$$

而 x_{16}, x_{19}任意。

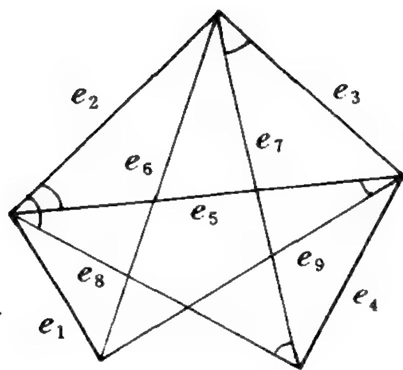
补充变数 $x_{58}, x_{67}, x_{59}, x_{78}, x_{69}$可自以下补充方程求得：

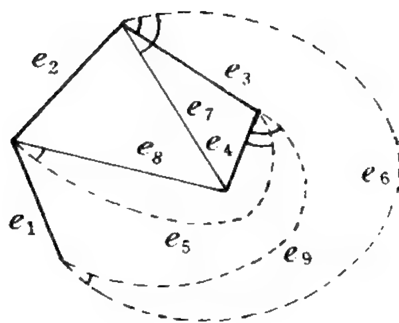
$$I_f(e_5, e_8) = 0 = x_{58} + x_{45},$$
$$I_f(e_6, e_7) = 0 = x_{67} + x_{36},$$
$$I_f(e_5, e_9) = 0 = x_{59} + x_{25},$$
$$I_f(e_7, e_8) = 0 = x_{78} + x_{37},$$
$$I_f(e_6, e_9) = 0 = x_{69} + x_{36} + x_{16} + x_{19}。$$

取 $x_{16} = x_{19} = 0$，则由此所得两扩充解列表如下：

	基本变数						补充变数				
	x_{25}	x_{28}	x_{36}	x_{37}	x_{45}	x_{49}	x_{58}	x_{67}	x_{59}	x_{78}	x_{69}
第一解	0	0	1	0	1	1	1	1	0	0	1
第二解	1	1	0	1	0	0	0	0	1	1	0

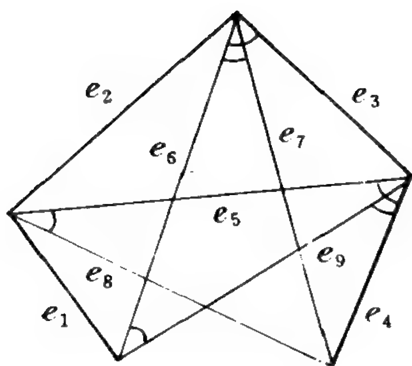
依据两解经合格化调整所得图形与原图比较，各如下所示：

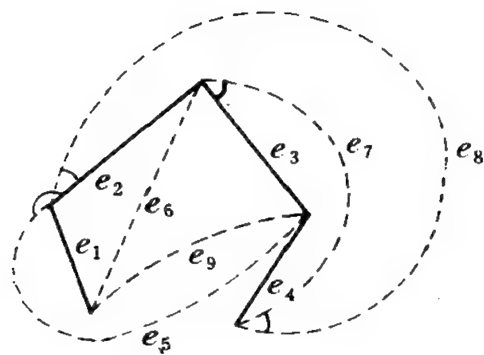




（第一解）

图 18





（第二解）

图 19

例 3　考虑线性图 G，最大树 T，树根 O 与 T－浸入 f 如图：

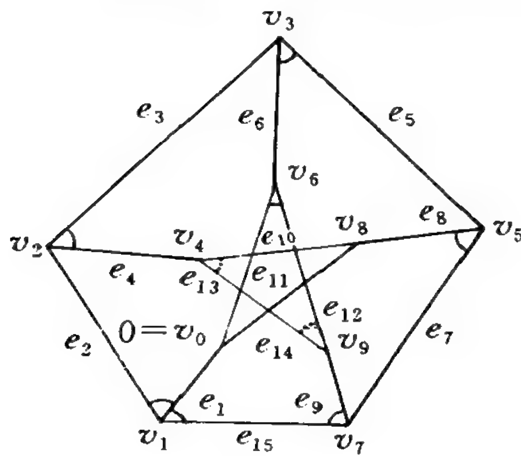


图 20

所处顶点	v_1	v_2	v_3	v_4	v_5	v_6	v_7	v_8	v_9	$O=v_0$
主要变数	$x_{2,15}$	x_{34}	x_{56}		x_{78}		$x_{9,15}$			$x_{1,11}$ $x_{1,14}$
补充变数				$x_{10,13}$		$x_{11,12}$		$x_{10,14}$	$x_{12,13}$	$x_{11,14}$

基本方程

$$I_f\ (e_{10},\ e_{11})\ =1=x_{34}+x_{56},$$
$$I_f\ (e_{10},\ e_{12})\ =1=x_{56}+x_{78},$$
$$I_f\ (e_{11},\ e_{13})\ =1=x_{34}+x_{56},$$
$$I_f\ (e_{12},\ e_{14})\ =1=x_{56}+x_{78},$$
$$I_f\ (e_{13},\ e_{14})\ =1=x_{34}+x_{78}。$$

由此知 e_{14}应从 G 中除去，而最后两方程可以略去。

$$I_f\ (e_{11},\ e_{15})\ =0=x_{56}+x_{2,15},$$
$$I_f\ (e_{14},\ e_{15})\ =0=x_{78}+x_{2,15},$$

由此知 e_{15}应从 G 中除去。

命除去 e_{14}，e_{15}后的图形为 G'。此时主要变数只有 x_{34}，x_{56}，x_{78}以及 $x_{1,11}$。基本方程只有两个：

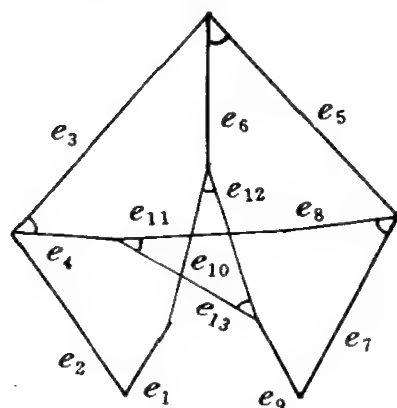
$$\begin{cases}x_{34}+x_{56}=1,\\ x_{56}+x_{78}=1。\end{cases}$$

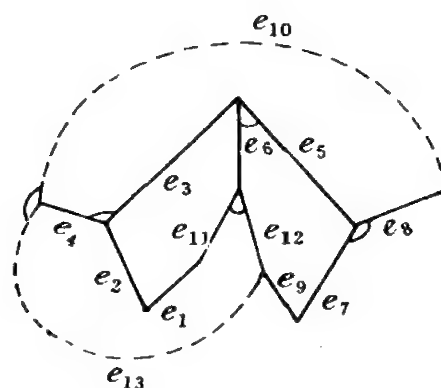
此时补充变数也只有三个：$x_{10,13}$，$x_{11,12}$，$x_{12,13}$。补充方程为

$$\begin{cases}I_f\ (e_{10},\ e_{13})\ =0=x_{10,13}+x_{78},\\ I_f\ (e_{11},\ e_{12})\ =0=x_{11,12}+x_{56},\\ I_f\ (e_{12},\ e_{13})\ =0=x_{12,13}+x_{56}。\end{cases}$$

	主　要　变　数			补　充　变　数		
	x_{34}	x_{56}	x_{18}	$x_{10,13}$	$x_{11,12}$	$x_{12,13}$
第一解	1	0	1	1	0	0
第二解	0	1	0	0	1	1

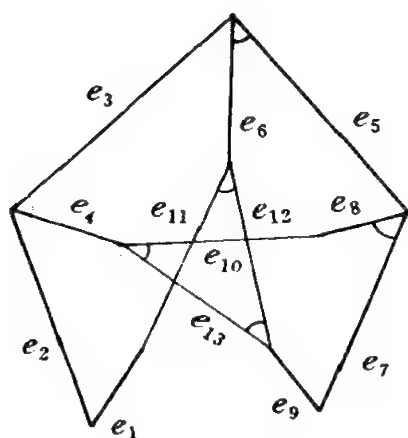
（$x_{1,11}$任意，可取为 0）

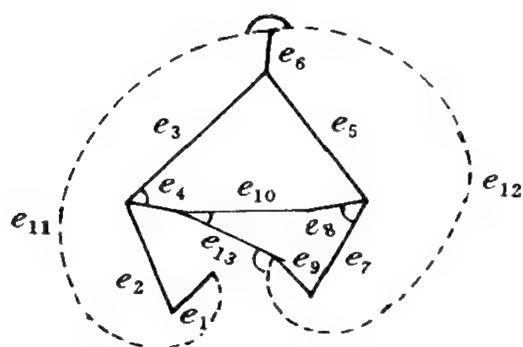




第一解

图 21





第二解

图 22

参 考 资 料

[1] L. Auslander. S. V. Parter, On imbedding graphs in the sphere, J.

Math. Mech., 10 (1961), 517—523.

[2] W. Bader, Topologische Problem der gedruckten Schaltung und seine Lösung, Archiv für Elektrotechnik, 49 (1964) 2—12.

[3] G. J. Fisher - O. Wing, Computer recognition and extraction of planar graphs from the incidence matrix, IEEE: Trans. on Cricuit Theory, CT - 13 (1966) 154—163.

[4] C. Kuratowski, Sur le probleme des courbes gauches en toploogie, Fund. Math. 15 (1930), 271—283.

[5] S. MacLane, A combinatorial condition for planar graphs, Fund. Math. 28 (1937). 22—32.

[6] H. Whitney, Non - separable and planar graphs, Trans. Amer. Math. Soc. 34 (1932), 339—362.

[7] 吴文俊，复合形在欧氏空间中的实现问题，Ⅰ，Ⅲ，数学学报，5 (1955)，505—552；8 (1958)，79—94。

拓扑中的量度与能计算性

恩格斯说过，数学研究现实世界中的空间形式与数量关系。空间形式与数量关系这两基本范畴之间，则往往通过量度概念联系起来。曲线的长度，多边形的面积，立体的体积，点集的测度，都是这种量度的具体实例。在历史上，量度往往指空间形式的某种度量性质。因之，作为以拓扑空间这种空间形式某些非度量性质的定性研究为主要任务的拓扑学，表面看来似乎与量度概念无关，但是，数量关系并非是一个单独或若干个数，而可广义地理解为某种数系统（或代数系统），如群、环、模、代数等等。因而，在这种广义的理解下，拓扑学中就有着极为丰富的量度。举例来说，拓扑空间的 0 维 Betti 数是一种量度，它代表空间连通分支的个数，1 维 Betti 数也是一种量度，它代表空间在某种意义下相互独立的闭圈个数，某种类型拓扑空间的尤拉数又是一种量度，它在拓扑学中的作用是颇为基本的。不仅这些个别的数可以看作拓扑空间的量度，而且在拓扑学发展过程中，还出现了各种以代数系统表达的各种量度：同调群或环，同伦群，K 环，以及最近出现的极小模型或 I^*—量度等，这些量度都反映了拓扑空间一定的拓扑性质。代数拓扑学正是由于这些量度的引入而蓬勃发展起来，而且还解决了不少数学其他领域中的难题，使代数拓扑学从无到有，成为 20 世纪中令人瞩目的一个大学科。试举几个简单实例来说明量度在拓扑学中的作用。

例 1 设 D^n 是 n 维单位球体，S^{n-1}是它的 $n-1$ 维球面，

f 是 D^n 到它自身的连续映像。Bronwer 一个著名的不动点定理是说：f 必有一不动点，即必有一点 x，使 $f(x)=x$。Bronwer 定理有许多不同的证明，它们往往需要一些巧思，依据量度的概念，则可提供一个较简单的证明如下。

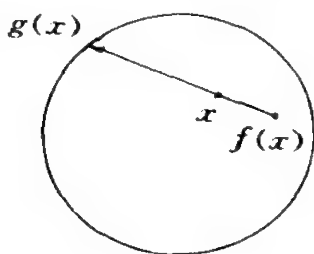


图 1

我们用反证法。假设 f 无不动点，则对任一 $x\in D^n$，从 $f(x)$ 到 x 的射线将与 S^{n-1} 交于一点 $g(x)$，于是 g（显然与 f 一样是连续的）有下面的性质：如果 $x\in S^{n-1}$，则 $g(x)=x$。现在假设有一种拓扑空间的量度 M，它是一种代数系统，具有以下一些性质（拓扑空间假定属于某种固定的类型，映像指连续映像）：

（1）对拓扑空间 X，Y 与映像 f：$X\to Y$，引出一代数同态 f_*：$M(X)\to M(Y)$（或 f^*：$M(Y)\to M(X)$）。

（2）对拓扑空间 X,Y,Z 与映像 f：$X\to Y$，g：$Y\to Z$，下面的代数同态图是可交换的，即 $(gf)_*=g_*f_*$（或 $(gf)^*=f^*g^*$）：

$$\begin{array}{ccc} M(X) & \xrightarrow{(gf)_*} & M(Z) \\ & {\scriptstyle f_*}\searrow \quad \nearrow{\scriptstyle g_*} & \\ & M(Y) & \end{array} \quad \text{或} \quad \begin{array}{ccc} M(X) & \xleftarrow{(gf)^*} & M(Z) \\ & {\scriptstyle f^*}\nwarrow \quad \swarrow{\scriptstyle g^*} & \\ & M(Y) & \end{array}$$

(3)若 id 是拓扑空间 X 到 X 的恒同映像，则 $(id)_*:M(x)\to M(X)$(或 $id)^*:M(X)\to M(X)$是 $M(X)$的恒同同态。

(4)$M(D^n)=0$，而 $M(S^{n-1})\neq 0$。

回到原来的 Bronwer 定理，记 S^{n-1}到 D^l 的嵌入映像为 $f_0 S^{n-1} \to D^n$，则 $gf_0: S^{n-1} \to S^{n-1}$即恒同映像，因而依据(1)、(2)有可交换图：

$$\begin{array}{ccc} M(S^{n-1}) & \xrightarrow{(id)_*} & M(S^{n-1}) \\ & {\scriptstyle f_{0*}} \searrow \quad M(D^n) \quad \nearrow {\scriptstyle g_*} & \end{array} \quad \text{或} \quad \begin{array}{ccc} M(S^{n-1}) & \xleftarrow{(id)^*} & M(S^{n-1}) \\ & {\scriptstyle f_0^*} \nwarrow \quad M(D^n) \quad \swarrow {\scriptstyle g^*} & \end{array}$$

但这些图显然与（3)、（4）冲突，由此知 f 无不动点的假设不成立，因而 Bronwer 定理得证。

任何一种具有性质（1）—（4）的量度都可用来证明 Bronwer 定理，这样的量度有 H_{n-1}，H^{n-1}（下或上同调群，系数群任意），π_{n-1}（同伦群）等。

例 2 设线性图 G，G 是否能嵌入平面是在理论上与实际上都极为重要的问题。1930 年以来 Kuratowski，Whitney，MacLane 等都给出了 G 能嵌入平面的充要条件，但这些条件都只是从问题的这一形式转变为另一形式，而且是“定性”的，要据之以检查一个具体的图是否能嵌入平面是困难的。我们则应用拓扑的某些量度给出了平面嵌入的“定量”的充要条件如下：从 G 作拓扑积 $G \times G$，在其中除去 σ_1，σ_2 有公共顶点的那些胞腔 $\sigma_1 \times \sigma_2$（σ_1，σ_2 是 G 的顶点或棱)，对余下的部分恒同 $\sigma_1 \times \sigma_2$ 与 $\sigma_2 \times \sigma_1$ 为 $\sigma_1 * \sigma_2 = \sigma_2 * \sigma_1$（$\sigma_1$，$\sigma_2$ 无公共顶点)，称所得复形为 $G * G$，于是 H^2（$G * G$，Z_2）可视为 G 的一种量度。在其中还可挑出一个特定的元素 Φ^2，它也是 G 的一种量度。我们的定理是

$$\Phi \text{ 可嵌入平面} \Leftrightarrow \Phi^2 = 0。$$

对于任意一个线性图 G，Φ^2 可通过具体计算知道它是否 $=0$，因而可由计算得知 G 是否能嵌入平面。这给出了这一问题解答的有效方法。

上面的条件还可进一步简化成下面的形式：从 G 可得出模 2 系数的线性方程组，它的有解与否是 G 能否嵌入平面的充要条件。

任意给一个线性图 G，G 能嵌入平面的情况是绝无仅有的。依据上面的线性方程组，我们容易得出应在 G 中丢弃哪一些棱，使余下的图 G' 能嵌入平面，而且还可据以指出 G' 嵌入平面的具体做法，以及不同嵌入的种数。这些步骤都是算法的，不难在计算机上予以实现。

上面两个简单实例说明，量度在现代拓扑学中的作用。但是，要使某种量度在应用上真正有效，必须具备若干条件。首先是不变性，应用于拓扑的量度必须是拓扑不变的，或者对所考虑的对象与问题要求同伦不变、组合不变、微拓不变等，正象古典几何中要求量度必须是量度不变或投影不变那样，否则这种量度将在根本上失去意义。其次是有限性，即量度能用例如有限个基所生成等，这样才适于掌握处理。因而，一个拓扑空间的奇异链群或一个微分流形的 de Rham 代数，虽然都是具有不变性的量度，由于它们超限生成而只能在推理过程中使用，但在定理的假设与终结中出现是不适宜的。除了这些隐含的要求之外，我们还提出：量度最好具有“能计算性”。这一概念可解释如下：在拓扑学中，经常要从若干给定的空间作出另一些新空间，例如从两个空间作拓扑积，沿某一部分粘合成拓扑和，作锥形，作吊形，作道路空间，作纤维丛等等。对于某种给定的量度 M，经常要从所给空间 X_1，X_2，……的量度 $M(X_1)$，$M(X_2)$，……定出所作新空间 Z 的量度 $M(Z)$，如果 $M(Z)$ 可以完全由 $M(X_1)$，$M(X_2)$，……以及反映这些空间 X_i 相互间几何关系的 $M(X_i)$ 的代数关系所完全决定，我们就说量度 M 对于所考虑的几何作法是“能计算的”，否则是“不能计算的”。

不言而喻，能计算的量度将给解决问题带来极大的方便，因而在应用上会有较大的威力。遗憾的是，在拓扑中广泛使用的种种量度，如同调，同伦等群或环，即使对极为简单的几何作法，也是“不能计算的”。例如：

（1）对于空间的拓扑积作法，实系数上同调环 $H_R{}^{**}$ 与整系数上同调群 $H_Z{}^*$ 都是能计算的，但整系数上同调环 $H_Z{}^{**}$ 则是不能计算的*。

（2）对于两个空间沿一公共子空间的粘合作法，实系数的上同调群 $H_K{}^*$ 是能计算的，但实系数上同调环 $H_K{}^{**}$ 或整系数上同调群 $H_Z{}^*$ 都是不能计算的。

以上（1）、（2）中能计算的部分各相当于 Künneth 公式与 Mayer－Vietoris 的序列，其中不能计算的部分则由简单实例可以证实。

拓扑学中的许多推理论证往往是迂回曲折的，其原因之一可能是拓扑中的主要工具——经典的同调，同伦等量度都往往是不能计算的。因而在推理过程中对于新出现空间的量度无法确定，而只能获得部分信息，而且还需要运用巧思走着迂回曲折的道路才能获得。这固然使拓扑学增添了引人入胜的色彩，但也增加了拓扑学研究的困难，为更高的发展添置障碍。

在 70 年代时出现了一种新的量度，它与经典量度的不同之处，是它具有高度的能计算性。

这一量度我们记之为 I^*，乃是美国拓扑学家 D. Sullivan 所引进。对于任一复形 K，可与微分流形那样构造外微分形式所定的代数 A^*（K）。这里 A^* 是一种量度，它是具有微分运算的实域上分次代数（简称 DGA），我们称之为 de Rham－

---

* 与传统的记法相异，我们把上同调 H 中的系数群或环写在 H 的右上角，以 * 表群而以 * * 表环结构。

Sullivan 代数。由于 A^* 是超限的，因而它本身是没有多大独立的意义的。但是在 K 有某些限制，例如单连通以及每一维数的 $H_R{}^*$ 都是有限生成时，Sullivan 从 A^* 纯代数地作出它的极小模型 $\mathrm{Min}A^*$，记之为 I^*，则 I^* 在每一维数都将有限生成而满足所谓有限性的条件。这一量度 I^* 又是同伦不变的，它包含了同伦同调中无限部分的全部信息。除此以外，我们从能计算的角度出发，对各种通常的几何作法证明了 I^* 量度都是能计算的，并给出了计算的具体表达式。在应用上，我们证明了不少过去用谱序列经过迂回曲折的推理在极强的条件下才能获得的结果，应用 I^* 量度可以在很一般的情况下得之甚易。

自然，我们对于 I^* 量度能计算性的研究局限于空间是单连通的情形，把它推广到较一般如所谓 nilpotent 空间应该不会是很难的，但我们没有这样做，因为这种推广是否必要，是否值得，是可以商榷的。真正的重要问题应该是：I^* 量度虽然是包含了同伦同调的无限部分，但由于它是实域上的 DGA，因而同伦同调的有限部分无法在 I^* 中得到反映。然而同伦同调的有限部分如挠系数之类都是应用上极其有力、成效卓著的武器，拓扑许多精彩的应用往往是通过它们取得的。如何推广 I^* 量度以包括同伦同调的有限部分而仍保有高度的能计算性，乃是一个有重大意义的课题。

自然，问题是十分困难的，虽然并不是没有线索可以入手。如果这个问题能够解决，则这一新的更广的 I^* 量度将取所有同伦同调的量度而代之，它可能使拓扑学的发展重点以及方法方向都为之改观。

博奕论杂谈：（一）二人博奕*

正象 17 世纪时概率论的产生与一些赌博问题有关那样，在本世纪发展起来的博奕论也与一些赌博以及下棋中的数学问题有关。1921 年法国的 E. Borel 为了用数学方法处理赌博一类问题，提出了“策略”这样一个概念，赌徒智力的高下就体现在是否能善于选择“策略”这一点上，这可以说是博奕论的萌芽。

我们先来举一个用扑克牌打赌的例子。甲乙两人各从一付扑克牌中选取 5 张后同时下注，赌注限定是 a 元或 b 元，此处 $a>b>0$，如果甲乙都下大注（a 元）或都下小注（b 元），则彼此摊牌，牌小者如注数付给牌大者（我们知道任意 5 张扑克牌组成的两组可以比较大小）；牌相同则不分胜负。如果甲下大注而乙下小注，则乙要求可以摊牌或认输。在乙认输时乙就付给甲 b 元，若乙要求摊牌，则乙牌小时须付甲 a 元，乙牌大时甲须付乙 a 元，牌相同时不分胜负，在甲下小注而乙下大注时也同样。

我们先问甲乙双方究有多少种不同的策略可取？

甲方取了牌以后，可以下大注，也可以下小注。如果甲下小注而发现对方下大注时，甲可以要求摊牌或是认输，因此总起来有三种可能的情形。所谓甲的一个“策略”，乃是甲所有可能发生的情况都决定了如何应付的一个全盘的方针。也就是

---

* 本文摘自《数学通报》，1959 年 10 月。

对每 5 张选取的牌都必须考虑到三种应付方法的一种，因为从 52 张扑克牌中选取 5 张共有 $C^5_{52}=2，598，960$ 种方法，所以甲的策略总数 $=3^{2,598,960}$。乙的策略总数也是一样。我们的问题是，甲乙在这许多策略中如何选出“最好”的策略来。

试再以下棋为例，所谓奕者的一个策略，乃是指从局始到局终就各种可能出现的情况都决定了应付之道的一个全盘的方针。自然，在所有常见的棋中，这样的策略的总数都是一个象恒河沙那样的大数，但总是一个有限的整数。所谓下棋下得高，就是善于在那恒河沙数的策略中尽可能选出较好的策略来。

虽然我们举的两个例子一“博”一“奕”，在表面上似乎至多只能是游戏一类的数学问题，可是就在那些例子里，就有一些因素是值得注意的：不论是博是奕，都有对立的两方进行竞赛，这种竞赛不仅有竞而且有争，这种对立的竞争，不仅出现在博奕中，在实际生活（例如体育竞赛），军事作战中，也都处处可见，博奕只是提供了这方面较显著也较简单因之较易分析掌握的实例，成为现在所谓博奕论思考问题时简单而具体的模型而已。事实上，就在 E. Borel 在 1921 年及其以后的一些博奕早期著作中，就已指出这方面的理论可能在商业，经济甚至军事上开阔应用数学方法的新途径。E. Borel 还就在这方面举了一些例子，下面是其中的一个：

有甲乙两支军队，在三个战场上作战，每个战场上双方兵数相同，相持不下，此外甲乙都拥有 15 个团的兵力作为后备部队。问题是甲乙将后备部队向各个战场增援时，应如何分配兵力（以团为单位）可使在尽可能多的战场上取得优势，战胜敌人。在这个问题中，甲乙有多少策略可取，应用组合排列是很容易算出的。

在已举的三个例子中，如果甲乙都取定了策略，不论是好

是坏，那么在每一个具体的“博奕”过程中，胜负的情况就完全确定了，这个胜负的情况，在简单的场合，可以用一个数表示出来，例如在扑克一例中，可用乙应付甲的元数来表示。在下棋一例中，不妨在甲胜时以 $+1$ 表示，甲负时以 -1 表示，和棋则以 0 表示，$+1$，-1 与 0 三数也可以看做乙应付给甲的数，虽然这些数的规定是有些任意的。在军队作战的例中，也可以用一个数来表示胜负程度，这样做虽然把问题简单化了，但对于问题的实质是并无损害的。

综合上述，我们可以提出这样一个“博奕”问题：

甲乙双方进行某种“博奕”，其中甲有 n 种策略可取，命之为 ξ_1，$\cdots$，ξ_n，乙有 m 种策略可取，命之为 η_1，$\cdots$，η_m。在甲取策略 ξ_i，乙取策略 η_j 时，乙应付与甲 a_{ij}。问题是：甲乙应取何种策略，对甲乙说来都可以认为“最好”？

试先就甲方来考虑这一问题，设甲先取策略 ξ_i，乙方必取一策略使乙付与甲的值为最小，因之甲在这时只希望获得 $\min\limits_j a_{ij}$。因之，就甲方而言，他应选取这样的一个策略 ξ_i，使相应的 $\min\limits_j a_{ij}$ 有最大值。这时甲所能从乙方取得的值为 $\max\limits_i \min\limits_j a_{ij}=V_{甲}$。对甲最合适的策略为符合条件 $\min\limits_j a_{ij}=V_{甲}$ 的策略 ξ_i，暂时可称为甲的最好策略。

次就乙方考虑，设乙选取策略 η_j，则甲必将取一策略，使甲从乙取得的值为最大，因之乙在此时须付与甲 $\max\limits_i a_{ij}$，故就乙方而言，他应选取这样的一个策略 η_j 使相应的 $\max\limits_i a_{ij}$ 有最小值，这时他付与甲方的值为 $\min\limits_i \max\limits_i a_{ij}=V_{乙}$，对乙最合适的策略为符合条件 $\max\limits_i a_{ij}=V_{乙}$ 的策略 η_i，暂时可称为乙的最好策略。

因为对任意 ξ_k，η_l 有

$$\min_j a_{ij}\leqslant a_{kl}\leqslant\max_i a_{ij},$$

故有

$$\max_i \min_j a_{ij} \leqslant \min_j \max_i a_{ij},$$

或即

$$V_{甲} \leqslant V_{乙}。$$

今先设 $V_{甲} = V_{乙}$。

此时设甲乙各自选取上面所说他们所认为的最好策略，设为 ξ_k 与 η_l，则应有

$$V_{甲} = \min_j a_{kj} \leqslant a_{kl},$$

$$V_{乙} = \max_i a_{il} \geqslant a_{kl},$$

合之得

$$V_{甲} \leqslant a_{kl} \leqslant V_{乙}。$$

由 $V_{甲} = V_{乙}$ 得

$$V_{甲} = V_{乙} = a_{kl}。$$

故乙方应支付甲方的值 a_{kl}，在甲方说来这是他所能期望得到的最大值。在乙方说来，也是他所期望支出的最小值，故 ξ_k，η_l 确合于策略“最好”的要求。

但一般来说，亦可有

$$V_{甲} < V_{乙}。$$

这时由 $V_{甲} \leqslant a_{kl} \leqslant V_{乙}$，

知至少有 $a_{kl} > V_{甲}$，

或

$$a_{kl} < V_{乙}。$$

在前一情形，乙所支付与甲的值 a_{kl} 将较甲所预期获得的最大值 $V_{甲}$ 为大，换言之，所谓乙的“最好”策略 η_1 实际上对乙说来并非“最好”。同样，在后一情形 $a_{kl} < V_{乙}$ 时，甲从乙所取得的值 a_{kl} 将较乙所预期支付与甲的值 $V_{乙}$ 为小，换言之，所谓甲的“最好”策略 ξ_k，实际上对甲说来并非“最好”。

由此可知，在 $V_{甲} = V_{乙}$ 时，上述博奕是完全确定的，甲乙双方各知应如何选择最好的策略以进行这一博奕，但在相反情形则甲乙将有不知应如何选择策略始为适当之感。

为了解除这一困难，E. Borel 与 Von Neumann 引进了“混合策略”的概念，而进行如下考虑：

先就甲方来说，甲方可以不简单地从可能策略 $\xi_1, \cdots, \xi_n$ 中选取其一，而可以概率 x_1 选取 $\xi_1, \cdots$，概率 x_n 选取 ξ_n，此处 $x_i \geqslant 0$，$\sum_{i=1}^{n} x_i = 1$。例如当同一博奕进行无数次时，$x_1, \cdots, x_n$ 各可视为选取策略 $\xi_1, \cdots, \xi_n$ 的频率，如果只就一次博奕而言，则 $x_1, \cdots, x_n$ 可表示甲对各该策略的偏爱或重视程度。

同样，乙方也可以概率 y_1 选取 η_1，…，概率 y_m 选取 η_m，此处 $y_i \geqslant 0$，$\sum_{j=1}^{m} y_i = 1$。

我们将称概率分布（x_1，…，x_n）为甲的一个混合策略，概率分布（y_1，…，y_m）为乙的一个混合策略，甲方一切可能的混合策略所成集合将记为 $s_{甲}$，同样有 $s_{乙}$。

于是问题成为：

甲乙应取何种混合策略，对甲乙说来都认为最好？

与前进行同样考虑如下：

设甲选取混合策略 $x = (x_1, \cdots, x_n)$，乙选取混合策略 $y = (y_1, \cdots, y_m)$，则乙应付与甲的数学期望显然应为

$$E(x, y) = \sum_{i,j} a_{ij} x_i y_j。$$

今就甲方来说，设甲先取混合策略 $x = (x_1, \cdots, x_n)$，乙必将先取一混合策略使乙付与甲的值（指数学期望）为最小，因之甲在此时只能希望获得 $\min\limits_{y \in s_{乙}} E(x, y)$，因之，就甲方而言，他应选取这样的一个混合策略 x，使相应的 $\min\limits_{y \in s_{乙}} E(x, y)$ 有最大值，这时他所能从乙方取得的值为

$$\max_{x \in s_{甲}} \min_{y \in s_{乙}} E(x, y) = V_{甲}。$$

满足
$$\min_{y \in s_{乙}} E(x, y) = V_{甲}$$

的混合策略 $x\in s_{甲}$ 将称为甲方的最好策略。

同样，就乙方来说，他所期望能付与甲的最小值为：

$$\min_{y\in s_{乙}}\max_{x\in s_{甲}} E(x, y)=V_{乙},$$

满足

$$\max_{x\in s_{甲}} E(x, y)=V_{乙}$$

的混合策略 $y\in s_{乙}$ 将称为乙方的最好策略。

与前同样应有

$$\max_{x\in s_{甲}}\ \min_{y\in s_{乙}} E(x, y)\leqslant\min_{y\in s_{乙}}\ \max_{x\in s_{甲}} E(x, y),$$

即

$$V_{甲}\leqslant V_{乙}。$$

在 $V_{甲}=V_{乙}$ 时，甲乙各可选择"最好"策略以进行博奕，若 $V_{甲}<V_{乙}$ 则否。

与前不同，在此时可证明恒有

$$V_{甲}=V_{乙}。$$

此即 Von Neumann 关于〈零和〉二人博奕的基本定理，今详细叙述此定理如下：

基本定理　设甲有 n 种策略可取：ξ_1，…，ξ_n。乙有 m 种策略可取：η_1，…，η_m。在甲取策略 ξ_i。乙取策略 η_j 时，乙应付与甲 a_{ij}；在甲取混合策略 $x=(x_1, x_2, \cdots, x_n)$，乙取混合策略 $y=(y_1, \cdots, y_m)$（此处 $x_i\geqslant 0$，$\sum x_i=1$，$y_i\geqslant 0$，$\sum y_j=1$ 或 $x\in s_{甲}$，$y\in s_{乙}$）时，乙应付与甲的数学期望为

$$E(x, y)=\sum_{\substack{1\leqslant i\leqslant n\\ 1\leqslant j\leqslant m}} a_{ij}x_iy_j, \tag{1}$$

则有

$$\max_{x\in s_{甲}}\ \min_{y\in s_{乙}} E(x, y)=\min_{y\in s_{乙}}\ \max_{x\in s_{甲}} E(x, y)。\tag{2}$$

注　这定理中甲乙开始时所能取的策略 ξ_1，…，ξ_n 与 η_1，…，η_m 各称为甲与乙的纯正策略。

这一定理是博奕论的基石，定理首先为 E. Borel 所提出，

混合策略的概念也是 E. Borel 所首先引入的。但 Borel 只能证明一些简单的特例。完全的证明首先为 J. von Neumann 所得，发表于 1928 年的 Mathematische Annalen 中，博奕论的诞生，可以说由这一篇短文开始。Von Neumann 并先后提供了四个方法各不相同的证明：

第一证明发表于 1928 年，依据实变函数论的一些简单知识；

第二证明发表于 1937 年，依据拓扑学中的 Bronwer 的定点定理；

第三证明发表于 1947 年，应用凸体理论的一些基本知识；

第四证明发表于 1950 年，应用了常微分方程。

其中第三证明最为简短与初浅，其前身是 J. Ville 应用凸体理论的一个初等证明（1938 年），其它三个证明都不能算是初等的。尤其是第一证明，所依据的虽只是实变函数论中一些最简单的概念：函数的连续性与半连续性，但推理颇为繁复，叙述冗长。虽然如此，第二证明系从第一证明演变而来，其次，第一证明的思路比较清楚，证明方法，也可以说是很自然的，与其它几个证明之尽管简短，巧妙，但颇为突兀者不同。最后，第一证明所证者实际上较上述基本定理所包含者为广，而这个内容较广的定理在有一些博奕论在经济上的应用中也是要用到的。因此下面将对这第一证明参照第二证明作一轮廓介绍，有些略去的论证补进去也并不困难，读者可自为之。

我们把欧氏空间中的一个点集中做凸的，如果对集中任意两点，连结它们的线段也全在这一点集里面。例如圆的内部与边界所成的点集，一个三角形或四面体的内部与边界所成的点集都是凸集。假设 x 是一个 n 维欧氏空间 R_n 中的一个凸集，Y 是一个 m 维欧氏空间 R_m 中的一个凸集，$f(x, y)$ 是定义在 $X \times Y$ 上的一个函数，即对每一 $x \in X$ 与 $y \in Y$，有一值

$f(x, y)$ 与之对应。我们说 $f(x, y)$ 是 $X \times Y$ 上的一个 K - 函数，如果对任意 $y \in Y$ 与任一实数 a，满足 $f(x, y) \geqslant a$ 的 X 中一切点 x 成一凸集，又对任意 $x \in X$ 与任一实数 a，满足 $f(x, y) \leqslant e$ 的 Y 中一切点 y 也成一凸集。容易看出，如果 R_n 的坐标系统是 $(x_1, \cdots, x_n)$，Y 中的坐标系统是 $(y_1, \cdots, y_m)$，那么 $X \times Y$ 上任一对 $x_1, \cdots, x_n$，而言是线性，又对 $(y_1, \cdots, y_m)$ 而言也是线性的所谓双线性函数 $f(x, y)$，必是 $\overline{X} \times \overline{Y}$ 上的一个连续的 K - 函数。于是 Von Neumann 所证明的内容较广的定理可述如下。

定理　如果 X，Y 各表欧氏空间 R_n 与 R_m 中的下述凸体，

$$\left.\begin{array}{l} X:\ x_1 \geqslant 0,\ \cdots,\ x_n \geqslant 0,\ x_1 + \cdots + x_n = 1, \\ Y:\ y_1 \geqslant 0,\ \cdots,\ y_m \geqslant 0,\ y_1 + \cdots + y_m = 1。\end{array}\right\} \tag{3}$$

其中 $(x_1, \cdots x_n)$ 与 $(y_1, \cdots, y_m)$ 各是 R_n 与 R_m 中的坐标；又 $f(x, y)$ 是 $X \times Y$ 上的任意连续 K - 函数，则必有

$$\operatorname*{Max}_{x \in X} \operatorname*{Min}_{y \in Y} f(x, y) = \operatorname*{Min}_{y \in Y} \operatorname*{Max}_{x \in X} f(x, y)。 \tag{4}$$

特别设 $f(x, y)$ 为 $X \times Y$ 上的双线性函数 $E(x, y) = \sum\limits_{i,j} a_{ij} x_i y_j$，即得博奕论的基本定理。

Von Neumann 的证明主要分成两步，先证明 $n = m = 2$ 时的特殊情形，这相当于博奕中甲乙双方都只有两个纯正策略可取的情形，其次再把一般的情形归结到这个特殊情形。

先考虑 $n = m = 2$ 的特殊情形。这时 X 由 $x_1 \geqslant 0$，$x_2 \geqslant 0$，$x_1 + x_2 = 1$ 所定。因此置 $x_1 = \xi$ 时，X 实际上相当于一个线段 $0 \leqslant \xi \leqslant 1$。同样置 $y_1 = \eta$，则 Y 相当于线段 $0 \leqslant \eta \leqslant 1$，而 $X \times Y$ 上的连续 K - 函数 $f(x, y)$，也就成为正方形 A：$0 \leqslant \xi \leqslant 1$，$0 \leqslant \eta \leqslant 1$ 上的一个连续 K - 函数 $f(\xi, \eta)$。由于函数 g 的连续性，对于每一个 ξ，$0 \leqslant \xi \leqslant 1$，使 $g(\xi, \eta) = \operatorname*{Min}_{0 \leqslant \eta \leqslant 1} g(\xi, \overline{\eta})$ 的值 η 必然存在，命这些 η 所成的点集为 K_ξ。又命

U 是正方形 A 中一切点（ξ, η）所成的点集，这里 $0\leqslant\xi\leqslant1$，而 $\eta\in K_\xi$。同样对任意 η，$0\leqslant\eta\leqslant1$，命满足 $g(\xi,\eta)=\max\limits_{0\leqslant\xi\leqslant1} g(\xi,\eta)$ 的一切值 ξ 的点集为 L_η（这样的 ξ 必存在）。又命 V 是正方形 A 中一切点（ξ, η）所成的点集，这里 $0\leqslant\eta\leqslant1$，而 $\xi\in L_\eta$。由于 g 是连续的 K－函数，可以证明 K_ξ，L_η 都是闭线段（也可以缩成一点）。又 U 与 V 都是“闭”点集，即 U 与 V 各含有它的一切极限点。

从直觉可以推测 U 与 V 应有公共交点（见附图）。事实上确实是如此。这一点可以直接从拓扑中的 Bronwer 定点定理导出，也可以根据实变函数论的一些简单事实来予以证明。在下面将应用 Bronwer 定理来证明它。

我们所要用到的 Bronwer 定点定理事实上只是这个定理的下述特殊情形；假设 h 是把一个正方形映到它自身的一个连续映象，换言之，对于正方形的每一点 z，有正方形中的一点 $h(z)$ 与之对应，而在点 z_1，z_2，…，z_n，…收敛于一点 z 时，$h(z_1)$，$h(z_2)$，…，$h(z_n)$，…也收敛于 $h(z)$。在这样的假定之下，h 必然有一个“定点”，即在正方形中有一点 z_0，使 $h(z_0)=z_0$。

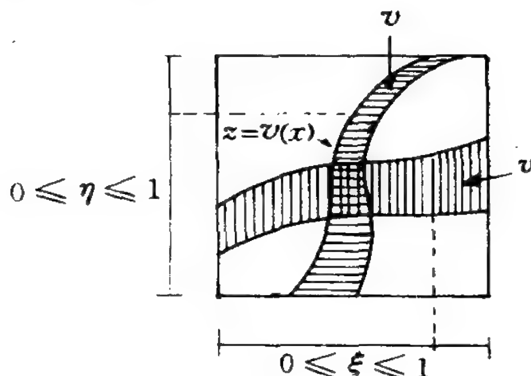


图 1

现在来证明 U 与 V 必有公共交点如下。显然在正方形 A 中但不在 U 中的点构成两个不相交开集 U_0 与 U_1 的和集，此

处 U_0 不含有边 $\eta=1$ 上的点，U_1 不含有 $\eta=0$ 上的点。同样在正方形 A 中但不在 V 中的点也构成两个不相交开集 V_0 与 V_1 的和集，此处 V_0 不含边 $\xi=1$ 上的点，V_1 不含边 $\xi=0$ 上的点。这里的 U_0，U_1，V_0，V_1 也可能是空集，但对我们的论证无关紧要。对于 A 中的任一点 z，命 $d(z, U)$ 与 $d(z, V)$ 各表 z 至点集 U 与 V 的距离，即 $d(z, U)$ [或 $d(z, V)$] 是 z 至 U（或 V）中各点的距离的下确界。由于 U 是闭集，$d(z, U)$ 同时也是 z 至 U 中各点的距离的最小值。当 $z\in U$ 时，$d(z, U)=0$，当 z 不$\in U$ 时，$d(z, U)>0$。对 $d(z, V)$ 来说也同样。今对正方形 A 中每一点 z 作一起点在 z 的一个矢量 $v(z)$，使它沿 ξ 与 η 方向的分量各为 $\pm d(z, U)$ 与 $\pm d(z, V)$，这里 $\pm$ 号的取舍如下：当 z 在 U_0（或 V_0）中时，$d(z, U)$（或 $d(z, V)$）前取 $+$ 号，当 z 在 U_1（或 V_1）中时，$d(z, U)$（或 $d(z, V)$）前取 $-$ 号。命矢量 $v(z)$ 的端点为 $h(z)$，则由关于 K_ξ，L_η 与 U，V 的已知性质，不难知 $h(z)$ 也在正方形 A 中，且 h 是一个连续映象。由前述 Bronwer 定理，在 A 中必有一点 z^*，使 $h(z^*)=z^*$，也就 $v(z^*)=0$，或 $d(z^*, U)=d(z^*, V)=0$。因之 $z^*\in U$ 也$\in V$，或 U，V 有公共交点 z^*。

记这一共公交点 z^* 为 (ξ^*, η^*)，则按 U 与 V 的定义有 $g(\xi^*, \eta^*)=\min\limits_{0\leqslant\bar{\eta}\leqslant1} g(\xi^*, \bar{\eta})$，$g(\xi^*, \eta^*)=\max\limits_{0\leqslant\bar{\xi}\leqslant1} g(\xi, \eta^*)$。由此得

$$\begin{aligned}&\max_{0\leqslant\xi\leqslant1}\min_{0\leqslant\eta\leqslant1} g(\xi, \eta)\geqslant\min_{0\leqslant\eta\leqslant1} g(\xi^*, \eta)\\&=g(\xi^*, \eta^*)=\max_{0\leqslant\xi\leqslant1} g(\xi, \eta^*)\\&\geqslant\min_{0\leqslant\eta\leqslant1}\max_{0\leqslant\xi\leqslant1} g(\xi, \eta)。\end{aligned}$$

另一方面易知有

$$\max_{0\leqslant\xi\leqslant 1}\min_{0\leqslant\eta\leqslant 1} g(\xi,\eta) \leqslant \min_{0\leqslant\eta\leqslant 1}\max_{0\leqslant\xi\leqslant 1} g(\xi,\eta)。$$

合之得

$$\max_{0\leqslant\xi\leqslant 1}\min_{0\leqslant\eta\leqslant 1} g(\xi,\eta) = \min_{0\leqslant\eta\leqslant 1}\max_{0\leqslant\xi\leqslant 1} g(\xi,\eta)。$$

这也就是前面的（4）式。这证明了 $n=m=2$ 时的情形。

在一般情形，由（3）式可将 $x_n=1-x_1-\cdots-x_{n-1}$ 与 $y_n=1-y_1-\cdots-y_{n-1}$ 代入 $f(x,y)$ 得一函数 $g(\overline{x},y)$，这里的 $\overline{x}=(x_1,\cdots,x_{n-1})$ 与 $\overline{y}=(y_1,\cdots,y_{n-1})$ 各展开与区域

$\overline{x}_1$：$\geqslant 0,\cdots,x_{n-1}\geqslant 0,x_1+\cdots+x_{n-1}\leqslant 1$，与 $\overline{Y}$：$y_1\geqslant 0,\cdots,y_{n-1}\geqslant 0,y_1+\cdots+y_{n-1}\leqslant 1$ 之上，所要证明的（4）式就变成

$$\max_{\overline{x}\in\overline{X}}\min_{\overline{y}\in\overline{Y}} g(\overline{x},\overline{y}) = \min_{\overline{y}\in\overline{Y}}\max_{\overline{x}\in\overline{X}} g(\overline{x},\overline{y}), \qquad (5)$$

其中 g 也是 $\overline{X}\times\overline{Y}$ 上的一个连续 K－函数。

在（5）式中求最大值与最小值的符号 $\max\limits_{\overline{x}\in\overline{X}}$ 与 $\min\limits_{\overline{y}\in\overline{Y}}$ 可以看做一个所谓运算子，其作用是把一个函数变成另一个函数，只是函数的变数减少了，就象偏微分 $\frac{\partial}{\partial x}$ 或积分 $\int_a^b\cdots dx$ 之类记号的作用一样。显然，在求 $\overline{x}\in\overline{X}$ 函数的最大值时，可依次对 x_{n-1}，x_{n-2}，$\cdots$，x_1 求最大值，换言之运算子 $\max\limits_{\overline{x}\in\overline{X}}$ 可以分解成若干个运算子的“积”。

$$\max_{\overline{x}\in\overline{X}} = \max_{x_1}\max_{x_2}\cdots\max_{x_{n-1}},$$

同样有

$$\min_{\overline{y}\in\overline{Y}} = \min_{y_1}\min_{y_2}\cdots\max_{y_{m-1}}。$$

容易证明把 Max_{xj} 或 Min_{yi} 作用在一个连续 K－函数上，所得的也是一个连续 K－函数，虽然所得的函数的变数已减少且定义域也已改变。根据已证明的特殊情形，作用于这函数上应

有

$$\max_{x_i}\min_{y_i}=\min_{y_i}\max_{x_i}。$$

（作用在一般的函数上时，这个式子是不对的）

在$\max\limits_{\bar{x}\in\bar{X}}\min\limits_{\bar{y}\in\bar{Y}}g(\bar{x},\ \bar{y})=\max_{x_1}\cdots\max_{x_{n-1}}\min_{y_1}\min_{y_2}\cdots\min_{y_{n-1}}g(\bar{x},\ \bar{y})$中，我们可将中间两运算子$\max_{x_{n-1}}$，$\min_{y_1}$交换（看做作用在函数$\min_{y_2}\cdots\min_{y_{n-1}}g(\bar{x},\ y)$上）；其次交换$\max_{x_{n-1}}$与$\min_{y_2}$，等等，最后得

$\min_{y_1}\cdots\min_{y_{n-1}}\max_{x_1}\cdots\max_{x_{n-1}}g(\bar{x},\ \bar{y})$。因之得（5）式，也就是（4）式。

至此定理已完全证明。

有了上面这一条定理，我们就可以引入这样一些概念：

（2）式左右两边的公共值v，我们叫做博奕的“值”。

矩阵（a_{ij}）叫做博奕的“支付函数”。

使$\min\limits_{y\in s_乙}E(x^*,\ y)=v$的$x^*\in s_甲$叫做甲的“最好策略”。

使$\max\limits_{x\in s_甲}E(x,\ y^*)=v$的$y^*\in s_乙$叫做乙的“最好策略”。

象前面所分析的那样，这里的最好策略的“最好”两字具有下述意义：甲取任一最好策略时，不论乙如何应付，甲可保证至少取得v（取得的值是指数学期望而言），若取别的策略，则不能保证取得v，同样，乙取任一最好策略时，不论甲如何应付，乙可保证至多付出v，若取别的策略，则不能保证只付出v。因此，这里的所谓最好策略，确实是当得起“最好”两字之称的。

在一个具体的博奕问题中，我们可以根据上述定理定出甲乙双方的最好策略来。

对前面所举扑克牌打赌的例来说，答案是相当繁复的。大体上说来，如果拿的牌相当好，就应该出大注，如果拿的牌不够好，就应按照一定的概率比例不规则地时而出大注，时而出小注，在出小注而遇见对方出大注时，就应立刻认输而从来不应要求摊牌。究竟好到什么程度才出大注，在不够好时出大注与小注的概率比例如何，都可以精确地计算出来。

在下棋例中，象象棋围棋，各人下子时对于两人过去所下的过程都是彼此清楚的，对这一类的棋，Zermelo 与 Von Neumann 曾经证明，两人都有确定的纯正策略为其最好策略。换句话说，未下之先，两人的胜负事实上早就已经判明了。虽然如此，爱下棋的同志并不必为此扫兴，因为为下棋的（纯正）策略有恒河沙数，Zermelo-Von Neumann 定理所证的只是在这恒河沙数的策略中每人有最好的策略存在，但要从恒河沙数的策略中真正的把它找出来，即使是最好的国手也是没法做到的。

对于军队在三个战场上作战的例子来说，可以证明，把全部援军开赴同一个战场，或者把援军平均分配于三个战场，都是拙劣的战略，应该予以摒弃。最值得考虑的战略是把全部兵力开赴两个战场，其分配是 10 个团与 5 个团，或 9 个团，与 6 个团，或 8 个团，与 7 个团，至于 11 个团与 4 个团的分配就是一个拙劣的战略了。其次值得考虑的战略是依下面的比例开赴三个战场：（9，5，1），（8，6，1），（7，7，1），也还有一些战略备考。对于各个不同战略考虑的程度，可以用概率具体表达出来。

在这里，请原谅作者忍不住要顺便提一下我国历史上一个著名的故事。在战国时，齐国的国王与齐国的大夫田忌赛马，齐王有三匹好马，上马中马下马。田忌也有三匹好马，上马中马下马。比赛分三次进行，每赛以千金作赌。齐王的马匹都个

个比田忌的好，虽然马力相差不多。看来田忌是输定了，可是田忌的门客孙膑（大家知道，这是中国历史上出色的大战略家，是传说孙子兵法的作者）给田忌出了一个主意：以下马对齐王的上马，而以上马对齐王的中马，中马对齐王的下马。结果田忌净胜千金。以这例与援军分配的例相较，虽然繁简有别，不同之处甚多，但，也有不少地方是类似的，就策略上来说，两者都应放弃一处，而争取在其它两处获胜。如果在赛马中只需要一些常识上的推敲，那么在援军分配问题中就不能不借助数学作出精确地判断了。

我们也可以应用博奕原理论来考虑诸如此类的许多问题：飞机对军舰的追逐，坦克的作战，等等。自然我们只能考虑简之又简的情况，要说真正能对实际问题发挥效果，而起指导作用，很可能还遥远得很，甚至还渺茫得很，只能寄希望于未来。但如果我们想到十七八世纪时代的概率论是怎样一个面目，而现代的概率论又怎样广泛地被应用于自然科学，工程技术，国民经济甚至日常生活之中。又如果我们想到博奕论所从事的对立现象正与概率论所从事的随机现象同样普遍地存在于现实世界，那么博奕论在将来的获得广泛应用，就不是一种想入非非的猜测了。事实上，博奕论的历史只有三十来年（从 Von Neumann1928 年的工作算起），可是已经出现了不少应用，特别值得一提的是与线性规划的联系以及在统计判决理论上的应用。Von Neumann 本人还曾应用之于工作的分配问题，生产系统按比例发展问题，与工厂建址问题，还以博奕论为基础发展了描述经济现象与社会现象的一个数学理论。虽然这些应用的深度与广度值得讨论，有的应用（特别在经济学上的应用）在观点上是根本有问题的，但已有的成果已足以鼓舞人们去作进一步的开拓。如果与概率作一比较，在十七八世纪的概率论仅能考虑一些类似排列组合的简单问题，直到 19 世纪下

半世纪才证明了大数定律并被成功地应用于气体分子理论与热力学，至于概率论基础的奠定，与随机过程理论的发展，还是近 30 年来的事，中间经历的路程是漫长而曲折的，因此才有 30 年历史的博奕论，内容不多就不足怪了。但是大风起于微末，谁又能武断目前还是微末的博奕论，不在将来蔚为大风呢？

数学的机械化*

十六、七世纪以来，人类历史上经历了一场史无前例的技术革命，出现了各种类型的机器，取代各种形式的体力劳动，使人类进入一个新时代。几百年后的今天，正如敬爱的周总理早在 1956 年就指出的那样，电子计算机已可开始有条件地代替一部分特定的脑力劳动，因而人类已面临另一场更宏伟的技术革命，处在又一个新时代的前夕。数学是一种典型的脑力劳动，它在这一场 新的技术革命中，无疑将扮演一个重要的角色。为了了解数学在当前这场革命中所扮演的角色，就应对机器的作用，以及作为数学的脑力劳动的方式，进行一定的分析。

一、什么是数学的机械化

不论是机器代替体力劳动，或是计算机代替某种脑力劳动，其所以成为可能，关键在于所需代替的劳动已经“机械化”，也就是说已实现了刻板化或规格化。正因为割麦、刈草、纺纱织布的动作已经是机械化刻板化了的，因而可据以造出割麦机、刈草机、纺纱机织布机来。也正因为加减乘除开方等运算这一类脑力劳动，几千年来就已经是机械地刻板地进行的，才有可能使得 17 世纪的法国数学家巴斯喀，利用齿轮传动造出了第一台机械计算机——加法机，并由莱布尼茨改进成为也

*本文摘自《百科知识》，1980 年。

能进行乘法的机器。数学问题的机械化，就要求在运算或证明过程中，每前进一步之后，都有一个确定的、必须选择的下一步，这样沿着一条有规律的、刻板的道路，一直达到结论。

在中小学数学的范围里，就有着不少已经机械化了的课题。除了四则、开方等运算外，解线性联立方程组就是一个很好的例子。在新编高中数学课本中，介绍了现代数学鼻祖高斯（德国人，1777～1855）解线性方程组的一种"消去法"，其求解过程是一个按一定程序进行的计算过程，也就是一种机械的、刻板的过程。根据这一过程编成程序，由电子计算机付诸实施，就可以不仅机器化而且达到自动化，在几分钟内求出一个未知数多至上百个的线性方程组的解答来，这在手工计算自然是不可能的。如果用手工计算，即使是解只有三、四个未知数的方程组，也将是繁琐而令人厌烦的。现代化的国防、经济建设中，大量出现的例如网络一类的问题，往往可归结为求解很多未知数的线性方程组。这使得已经机械化了的线性方程解法在四个现代化中起着一种重要作用。

即使是不专门研究数学的人们，也大都知道，数学的脑力劳动有两种主要形式：数值计算与定理证明（或许还应包括公式推导，但这终究是次要的）。著名的数理逻辑学家美国洛克菲勒大学教授王浩先生在一篇《向机械化数学前进》的有名文章中，曾列举了这两种数学脑力劳动的若干不同之点。我们可以简略而概括地把它们对比一下：

计算	证明
易	难
繁	简
刻板	灵活
枯燥	美妙

计算，如已经提到过的加减乘除开方与解线性方程组，其所以虽繁而易，根本原因正在于它已经机械化。而证明的巧而难，是大家都深有体会的，其根本原因也正在于它并没有机械化。例如，我们在中学初等几何定理的证明中，就经常要依靠诸如直观、洞察、经验，以及其他一些模糊不清的原则，去寻找捷径。

二、从证明的机械化到机器证明

一个值得提出的问题是：定理的证明是不是也能象计算那样机械化，因而把巧而难的证明，化为计算那样虽繁而易的劳动呢？事实上，这一证明机械化的设想，并不始自今日，它早就为 17 世纪时的大哲学家、大思想家和大数学家莱布尼茨所具有。只是直到 19 世纪末，希尔伯特（德国数学家，1862～1943）等创立并发展了数理逻辑以来，这一设想才有了明确的数学形式。又由于 40 年代电子计算机的出现，才使这一设想的实现有了现实可能性。

从本世纪二、三十年代以来，数理逻辑学家们对于定理证明机械化的可能性，进行了大量的理论探讨，他们的结果大都是否定的。例如哥德尔（Godel）等人的一条著名定理就说，即使看来最简单的初等数论这一范围，它的定理证明的机械化也是不可能的。另一面，1950 年波兰数学家塔斯基（Tarski）则证明了初等几何（以及初等代数）这一范围的定理证明，却是可以机械化的。只是塔斯基的结果近于例外，在初等几何及初等代数以外的大量结果都是反面的，即机械化是不可能的。

1956 年以来美国开始了利用电子计算机做证明定理的尝试。1959 年王浩先生设计了一个机械化方法，用计算机证明了罗素等著的《数学原理》这一经典著作中的几百条定理，只

用了9分钟，在数学与数理逻辑学界引起了轰动。有一时机器证明的前景似乎非常乐观。例如1958年时就有人曾经预测：在10年之内计算机将发现并证明一个重要的数学新定理。还有人认为，如果这样，则不仅许多著名哲学家与数学家如庇阿诺、怀特海、罗素、希尔伯特以及杜灵等人的梦想得以实现，而且计算将成为科学的皇后，人类的主人！

然而，事情的发展却并不如预期那样美好。尽管在1976年时，美国的哈肯等人，在高速计算机上用了1，200小时的计算时间，解决了数学家们100多年来所未能解决的一个著名难题——四色问题*，因此而轰动一时，但是，这只能说明计算机作为定理证明的辅助工具有着巨大潜力，还不能认为这样的证明就是一种真正的机器证明。用王浩先生的说法，哈肯等关于四色定理的证明是一种使用计算机的特例机证，它只适用于四色这一特殊的定理，这与所谓基础机器证明之能适用于一类定理者有别。后者才真正体现了机械化定理证明，进而实现机器证明的实质。另一面，在真正的机械化证明方面，虽然泰斯基在理论上早已证明了初等几何的定理证明是能机械化的，还提出了据以造判定机也即是证明机的设想，但实际上他们的机械化方法非常繁，繁到不可收拾，因而远远不是切实可行的。1976年时，美国做了许多在计算机上证明定理的实验，在泰斯基的初等几何范围内，用计算机所能证明的只是一些近于同义反复的“儿戏式”的“定理”。因此，有些专家曾经发出过这样悲观的论调：如果专依靠机器，则再过100年也未必能证明出多少有意义的新定理来。

---

* 1852年英国人格斯里克（Guthric）提出的猜测：“就地图着色而言，四色是足够的。也就是说，对于任何一幅地图着色，只需四种颜色，就足以使得所有相邻地区的颜色不致重复。

三、一条切实可行的道路

1976年冬，正值四人帮垮台之际，我们开始了定理证明机械化的研究。1977年春取得了初步成果，证明初等几何主要一类定理的证明可以机械化。在理论上说来，我们的结果已包括在泰斯基的定理之中。但与塔斯基的结果不同，我们的机械化方法是切实可行的，即使用手算，也可以证明一些艰深的定理。

我们的方法主要分两步，第一步是引进坐标，然后把需证定理中的假设与终结部分都用坐标间的代数关系来表示。我们所考虑的定理局限于这些代数关系都是多项式等式关系的范围，例如平行、垂直、相交、距离等关系都是如此。这一步可以叫做几何的代数化。第二步是通过代表假设的多项式关系把终结多项式中的坐标逐个消去，如果消去的结果为零，即表明定理正确，否则再作进一步检查。这一步完全是代数的，即用多项式的消元法来验证。

上述两步都可以机械与刻板地进行。根据我们的机械化方法编成程序，以在计算机上实现机器证明，并无实质上的困难。事实上数学所某些同志以及国外的王浩先生都曾在计算机上试行过。我们自己也曾在国产的长城203台式机上证明了像西姆森线* 那样不算简单的定理。1978年初，我们又证明了初等微分几何中主要的一类定理证明也可以机械化。而且这种机械化方法也是切实可行的，并据此用手算证明了不算简单的一些定理。

从我们的工作中可以看出，定理的机械化证明，往往极度

---

* 从圆周上任一点向圆内接三角形的三条边做垂线，三垂足必在一条直线上，这条直线叫西姆森线，这条定理叫西姆森定理。

繁复，与通常既简且妙的证明形成对照，这种以量的复杂来换取质的困难，正是利用计算机所需要的。

在电子计算机如此发展的今天，把我们的机械化方法在计算机上实现不仅不难，而且有一台微型的台式机也就够了。就像我们曾经使用过的长城203，它的存数最多只能到234个10进位的12位数，就已能用以证明西姆逊线那样的定理。目前内存简单的所谓个人用微型机，已到32K以至64K字节。稍高级的已到256K以至400多K字节，而且随着超大规模集成电路与其它技术的出现与改进，微型机将愈来愈小型化而内存却愈来愈大，功能愈来愈多，自动化的程度也愈来愈高。用不着到20世纪末，这一类方便的小型机器就可为广大群众普遍使用。它们不仅将成为证明一些不很简单的定理的武器，而且还可用以发现并证明一些艰深的定理，而这种定理的发现与证明，在数学研究手工业式的过去，将是不可想象的。

应该指出，目前我们所能证明的定理，局限于已经发现的机械化方法的范围，例如初等几何与初等微分几何之内。而如何超出与扩大这些机械化的范围，则是今后需要探索的长期的理论性工作。

四、历史的启示与未来的技术革命

作为结束，我们提出几点看法。

首先，成功的机械化方法并非始自这几年。约在一年以前，我们发现早在1899年出版的希尔伯特的经典名著《几何基础》中，就有着一条真正的正面的机械化定理：初等几何中只涉及从属与平行关系的定理证明可以机械化。当然，原来的叙述并不是以机械化的语言来表达的，也许就连希尔伯特本人也并没有对这一定理的机械化意义有明确的认识，自然更不见

得有其他人提到过这一定理的机械化内容。希尔伯特这一名著是以公理化的典范而著称于世的，但我认为，该书更重要之处，是在于提供了一条从公理化出发，通过代数化以到达机械化的道路。自然，处于希尔伯特以及其后数学的一张纸一支笔的手工作业时代里，公理化的思想与方法得到足够的重视与充分的发展，而机械化的方向与意义受到数学家的忽视是完全可以理解的。但在电子计算机已日益普及，因而繁琐而重复的大量计算已成为不足道的现代，机械化的思想应比公理化思想受到更大重视，似乎是合乎实际的。

其次应该着重指出，我们从事机械化定理证明工作获得成果之前，对泰斯基的已有工作并无接触，更没有想到希尔伯特的《几何基础》会与机械化有任何关系。我们是在中国古代数学的启发之下提出问题并想出解决办法来的。

说起来道理也很简单：中国的古代数学基本上是一种机械化的数学。四则运算与开方的机械化算法由来已久。汉初完成的《九章算术》中，对开平、立方的机械化过程，就有详细说明，到宋代更发展到高次代数方程求数值解的机械化算法。在《九章算术》中还有着各种线性联立方程组的问题与解法以及正负数的概念，在魏晋时刘徽的《九章算术》注中，说明了几种机械的消去法及其详细的机械化算法过程。把刘注的说明列成图表，即与前面所提到的中学课本中所列高斯消去法的那些图表无异，宋代秦九韶《数书九章》中，更有着颇为繁复的算题与详细图表。沈康身同志所著的《中国数学史略》，在这方面做了详细介绍。

在宋元时代，我国就创立了“天元术”，引进了天元，以及天元、地元、人元、物元等相当于现代未知数的概念，把许多问题特别是几何问题转化成代数方程与方程组的求解问题。这一方法用于几何可称为几何的代数化。12 世纪的刘益将新法与“古法”比较，称“省功数倍”。与之相伴而生，又引进

了相当于现代多项式的概念，建立了多项式的运算法则和消元法的有关代数工具，使几何代数化的方法得到了有系统的发展，具见于宋元时代幸以保存至今的杨辉、李冶、朱世杰的许多著作之中。几何的代数化是解析几何的前身，这些创造使我国古代数学达到了又一个高峰。可以说，当时我国已到达了解析几何与微积分的大门，具备了创立这些数学关键领域的条件，但是各种原因使我们数学的雄伟步伐就在这些大门之前停顿下来。几百年的停顿，使我们这个古代的数学大国在近代变成了数学上的纯粹入超国家。然而，我国古代机械化与代数化的光辉思想和伟大成就是无法磨灭的。作者本人关于数学机械化的研究工作，就是在这些思想与成就启发之下的产物，它是我国自《九章算术》以迄宋元时期数学的直接继承。

恩格斯曾经指出，枪炮的出现消除了体力上的差别，使中世纪的骑士阶级从此消声匿迹，为欧洲从封建时代进入到资本主义时代准备了条件。近年有些计算机科学家指出，个人用计算机的出现，其冲击作用可与枪炮的出现相比。枪炮使人们在体力上难分强弱，而个人用计算机将使人们在智力上难分聪明愚鲁。又有人对数学的未来提出看法，认为计算机的出现，将使数学现在一张纸一支笔的方法，在历史的长河中，无异于石器时代的工业方法。今天的数学家们，不得不面对计算机的挑战，但是，也不必妄自菲薄。大量繁复的事情交给计算机去做了，人脑将仍然从事富有创造性的劳动。

我国在体力劳动的机械化革命中曾经掉队，以致造成现在的落后状态。在当前新的一场脑力劳动的机械化革命中，我们不能重蹈覆辙。数学是一种典型的脑力劳动，它的机械化有着许多其它类型脑力劳动所不及的有利条件。周总理的遗愿，我国古代数学的光辉，都鼓舞着我们为实现数学的机械化，在某种意义上也可以说是真正的现代化而勇往直前。

几何定理的机器证明*

用机械方法证明定理的思想，可以远溯至 17 世纪的莱布尼茨，在本世纪内，已经由希尔伯特的数理逻辑学派和他的学生们用精确的数学形式表述出来。这问题的实质在于：把通常数学证明中所固有的质的困难性，代之以用算法方式使证明过程标准化而造成的计算中的量的复杂性。这种属于计算的量的复杂性过去是人力所不可及的，而现在由于计算机的产生和迅速发展，已经变得越来越微不足道了。但是，虽然经过了巨大的努力，这方面的研究却时常得到否定的结果，即形成不可判定的数学理论，要举一个著名的肯定的结果，我们可以提到塔斯基（Tarski）用机械方法证明初等几何和初等代数中的定理的方法。塔斯基的方法以及后来的一些方法在很大程度上是以斯图谟定理的推广为根据的，这些方法仍旧太复杂，即使用计算机也实现不了。关于在美国用计算机所作实验及所得结果，可参阅 IEEE's Trans. on Computers, 1976，机器证明定理专辑中的一些论文。

1977 年，作者发现了一种方法，它所根据的原理与塔尔斯基等人所用的完全不同，用了这种方法，可以用机械方式给出很难的几何定理的证明。后来这方法又被扩充到初等微分几何，我们用不同的计算机编了程序，作了试验。本文是关于我们所用方法和迄今为止用计算机所作试验的初步报告。

---

* 本文摘自《自然杂志》3 卷 12 期，1980 年。

现在从初等几何讲起。为举例起见，我们来考察费尔巴赫(Feuerbach)定理，这定理可叙述如下：

三角形的九点圆与它的 4 个内切和旁切圆相切。

取给定三角形的 3 个顶点为$(2u_i, 2u_{i+3})$, $i=1,2,3$, 九点圆的圆心和半径为(x_4, x_5)和 x_8, 内切圆或任一旁切圆的圆心和半径为(x_6, x_7)和 x_9。设 $2x_1, 2x_2, 2x_3$ 分别对应于三角形的 3 条边的长度，其值或正或负，依照所证情形是内切圆或旁切圆而定。

为简单起见，设*

$$u_{ij} = u_i - u_j,\ i, j = 1, \cdots, 6,$$

$$2\Delta = \begin{vmatrix} u_1 & u_4 & 1 \\ u_2 & u_5 & 1 \\ u_3 & u_6 & 1 \end{vmatrix},$$

$$\alpha_1 = u_{56}u_{64}u_{45} + u_1{}^2 u_{56} + u_2{}^2 u_{64} + u_3{}^2 u_{45} - 2(u_2u_3u_{56} + u_3u_1u_{64} + u_1u_2u_{45},$$

$$\beta_1 = -u_{23}u_{31}u_{12} - u_4{}^2 u_{23} - u_5{}^2 u_{31} - u_6{}^2 u_{12} + 2(u_5u_6u_{23} + u_6u_4u_{31} + u_4u_5u_{12}),$$

$$\alpha_2 = [u_{56}u_{64}u_{45} + u_4(u_2{}^2 - u_3{}^2) + u_5(u_3{}^2 - u_1{}^2) + u_6(u_1{}^2 - u_2{}^2)] + u_{56}x_2x_3 + u_{64}x_3x_1 + u_{45}x_1x_2,$$

$$\beta_2 = u_{23}u_{31}u_{12} + u_1(u_5{}^2 - u_6{}^2) + u_2(u_6{}^2 - u_4{}^2) + u_2(u_4{}^2 - u_5{}^2) + u_{23}x_2x_3 + u_{31}x_3x_1 + u_{12}x_1x_2,$$

---

* 下面的数据并不是用机械自动的手续得来，而是经过某些推理论证以及一些简化得来，因而从机器证明的角度来说并不完全合适。这是由于当时使用的计算机过小，自动化程度过低，作者对计算机又是初次使用，为了使计算量不致超出所用计算机的范围，采取了这样的办法。事实上，现在已可编成程序自动地把几何命题中的几何条件变成代数数据。这样费尔巴赫定理的机器证明从几何假设到证明完毕全部可由机器完成，所用数据自然与这里所举的不会一样。

$$\gamma=(u_{31}u_{12}+u_{64}u_{45})x_1+(u_{12}u_{23}+u_{45}u_{56})x_2+(u_{23}u_{31}+u_{56}u_{64})x_3+x_1x_2x_3。$$

于是费尔巴赫定理的假设等价于下列一组关系式$(I)_F$：

$$(I)_F\begin{cases}f_1\equiv x_1^2-u_{23}^2-u_{56}^2=0,\\ f_2\equiv x_2^2-u_{31}^2-u_{64}^2=0,\\ f_3\equiv x_3^2-u_{12}^2-u_{45}^2=0,\\ f_4\equiv 4\Delta x_4-a_1=0,\\ f_5\equiv 4\Delta x_5-\beta_1=0,\\ f_6\equiv 2\Delta x_6-\alpha_2=0,\\ f_7\equiv 2\Delta x_7-\beta_2=0,\\ f_8\equiv 4\Delta x_8-x_1x_2x_3=0,\\ f_9\equiv 2\Delta x_9+\gamma=0。\end{cases}$$

这定理的结论等价于下面的关系式

$$g\equiv(x_8-x_9)^2-(x_4-x_5)^2-(x_6-x_7)^2=0。$$

最后，这定理的真确性等价于说：当 $\Delta\neq 0$，即当三角形是非退化的，它的 3 个顶点不坍塌到位于一直线上时，$g=0$ 就是$(I)_F$ 的一个形式推论。

为了对上述定理作出机器证明，我们先指出如下几点：

(1) 变数 u_i 可当作参数，这相当于实际上在所有几何定理中通常不言而喻地假定的事实，即所涉及的三角形是一个一般的三角形，从而把 $\Delta=0$ 的退化情形排除在外。

(2) 变数 x_i 为 $(I)_F$ 中具有整系数的代数关系式所约束，并且每次一个地相继呈现。

(3) 方程组 $(I)_F$ 定义了一个有理数（或实数）K 域上实 6 维的代数簇 V_F。在这情形中，这个簇是不可约的，它具有超越次数为 6 的一般点（u_1，…，u_6，x_1，…，x_9），这里 u_i 是未定元，x_i 是代数的，由方程 $f_i=0$ 定义。

(4) 上述费尔巴赫定理的意义就是说在代数簇 V_F 上 $g=0$。

这个情况是很一般的，上列各点可以应用于初等几何中的大部分定理。

事实上，初等几何中凡是只包含希尔伯特公理系统中的关联公理、迭合公理和平行公理而不包含次序公理和连续公理的大多数定理，都可以表述成下面的形式。

所述定理的假设用下式给出（或易于变换成下式）：

$$\begin{cases} f_1 \equiv C_{10} x_1{}^{m_1} + C_{11} x_1{}^{m_1-1} + \cdots + C_{1m_1} = 0, \\ f_2 \equiv C_{20} x_2{}^{m_2} + C_{21} x_2{}^{m_2-1} + \cdots + C_{2m_2} = 0, \\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ f_N \equiv C_{N0} x_N{}^{m_N} + C_{N_1} x^{N^{m_N-1}} + \cdots + C_{N_{m_N}} = 0。 \end{cases} \qquad (\mathrm{I})$$

每一个 C_{ij} 是含有一些未定元或参数 $u_1, \cdots, u_d$ 以及变数 $x_1, \cdots, x_{i-1}(i=1, \cdots, N)$的多项式，具有有理系数或者甚至整系数。需要证明的结论是

$$g=0,$$

其中 g 是含有 u_i和 x_j 的多项式，具有有理系数或者甚至整系数。

于是这定理等价于说 $g=0$ 是在

$$C_{i0} \neq 0$$

的附加条件下关系式（I）的形式推论。这附加条件相当于这样的事实：我们只对非退化或一般情形感兴趣。问题是要设计出一种方法，使我们能用可行的机械方式判断这定理的真伪，所谓可行，就是说我们能用手和笔，或者借助于不太大的电子计算机，在不太长的时间内完成证明的步骤。

我们的方法进行如下：

将 g 依次除以 f_N，f_{N-1}，$\cdots$，f_1，它们依次被当作 x_N，

$x_{N-1}, \cdots, x_1$ 的多项式。设最后的余式是 R，使我们得到下面这样的恒等式（$S_i \geqslant 0$）：

$$C_{10}{}^{S_1} \cdots C_{r0}{}^{S_r} = \sum A_i f_i + R,$$

其中 A_i 和 R 是 u 和 x 的多项式，具有有理（或整）系数，在每一个 x_i 中 R 的次数 $< m_i$。把（I）当作基域 K 上代数簇 V 的定义方程，于是我们的方法就以下面这容易证明并且几乎是平凡的原理为基础：

原理 如果 V 在具有真正的实维 d（= 未定元 u_i 的数目）的 K 域中不可约（这具有明显的几何意义），那末作为多项式的 $R \equiv 0$ 是在 $C_{i0} \neq 0$ 的一般情形中定理为真的充要条件。在任何情况下，$R \equiv 0$ 是使一般情形中定理为真的充分条件。

我们应用了上述原理来证明初等几何中的各种定理，我自己用的是 HP9835A 台式计算机，陈东岳用的是 Felix $C-512$ 计算机。

在试验过的定理中可以举出这样一些：勾股定理（即毕达哥拉斯定理），一些三角恒等式，西姆森线（Simson line）定理，派泼斯（Pappus）定理，帕斯卡定理，费尔巴赫定理，等等，特别是对于费尔巴赫定理，陈用 Felix C 计算机以不到 12.5 分的时间完成了证明，所用语言 FCY 是通常用的 LISP 的改进，并且是陈自己研究成功的。

一旦定理的证明变得平凡时，我们就尝试去发现“新”定理。所谓“新”是指通常不知道或文献中不易找到，或者真正是完全新的。于是，按照同样的方式，我们用 HP9835A 台式计算机发现了并证明了两个这样的“新”定理。

为了说明我们发现的第一个“新”定理，让我们考察分别由 A_1，A_2，A_3 和 B_1，B_2，B_3 各 3 点组成的两个点集，它们各自位于同平面内的 L_A 和 L_B 两条线上。设 B 点按任何方式排列，例如 B_i，B_j，B_k。根据派泼斯定理，3 个交点 A_2B_k

$\wedge A_3B_j$，$A_3B_i \wedge A_1B_k$ 和 $A_1B_j \wedge A_2B_i$ 将位于同一条派泼斯线上，我们用 $[i, j, k]$ 来表示。在 B 点的不同排列之下，我们一共得到 6 条这样的派泼斯线，于是我们可以提出这样的问题：这 6 条线是否与同一个二次曲线相切？通过粗略地画图，我们看到很可能真是如此，不过二次曲线退化成两点。我们不用通常的方式去证明，因为这不是我们的目的，我们应用上述方法，借助 HP9835A 计算机。我们终于发现并证明了下述“新”定理：

3 条派泼斯线 [123]，[231] 和 [312] 是共点的。另外 3 条派泼斯线 [132]，[321] 和 [213] 也是如此。

关于第二个“新”定理，我们考察同一个二次曲线上的 6 个点 A_1，…，A_6。我们把从它们得到的任何交点 $A_iA_j \wedge A_kA_l$（i，j，k，l 都不相同）叫做帕斯卡点。这些帕斯卡点共有 45 个，根据帕斯卡定理，将会 3 个 3 个地位于总共有 60 条的所谓帕斯卡线上。于是我们可以提出如下的问题：在 45 个帕斯卡点中，怎样的 6 点组合会位于同一个二次曲线上？

为了这目的，让我们考察由 6 个数字 1，2，…，6 组成的对称群，以及譬如说由循环 $g=(1\quad 2\quad 3\quad 4\quad 5\quad 6)$ 产生的 6 阶子群。现在考察任何帕斯卡点 $P=A_iA_j \wedge A_kA_l$，我们来看一看，究竟 6 个帕斯卡点 g^iP，$i=0$，1，…，5 会不会位于同一个二次曲线。经过试验我们看到，除了 6 点将位于两条帕斯卡线（重合或不重合）的退化二次曲线的平凡情况外，唯一需要考虑的可能情况是 $P=A_1A_3 \wedge A_2A_5$ 或类似的情况。我们仍旧用 HP9835A 发现并证明了如下的“新”定理：

当 $P=A_1A_3 \wedge A_2A_5$ 时，6 个帕斯卡点 g^iP，$i=0$，1，…，5 将位于同一个二次曲线，叫做帕斯卡二次曲线。对于不同的 g 和不同的 P，我们共得到 20 个这样的帕斯卡二次曲线。

60 条帕斯卡线组成了一个很有趣的构形，这种构形曾在 19 世纪被一些几何学家象斯泰纳（Steiner），凯莱（Cayley），萨尔门（Salmon），寇克曼（Kirkmann）等作过很深的研究，例如可参阅萨尔门的 Conic Section 的附录。对于 20 个帕斯卡二次曲线的构形（或 60 条帕斯卡线一起），也可以作同样的研究。

用来发现和证明这两个"新"定理的 HP9835A 计算机是台式计算机，存储量是 256 K 个字节，其中实际用到的只有 100K 个字节。它的速度很低，大约每秒 3000 次加法。证明帕斯卡二次曲线定理所用时间是 60 小时不到一点。如果用速度较高的计算机，譬如说陈所用的每秒 370000 次加法的 Felix C 计算机，所需时间可以缩短到 30 分左右。根据所有这些情况，我们认为我们的方法是很有效的，它使我们可以运用小型中速计算机（由于今天的技术发展程度，这是很平常的）去发现和证明比已知定理复杂得多的"新"定理，这些定理用通常的方法也许是完全不可能证明的。

上述方法适当作些改变后也可以用于微分几何。

数学的机械化问题*

通常所谓的机械化，不论是工业的还是农业的，都是指体力劳动的机械化而言。现在我们提出这样的问题：数学研究这样的脑力劳动，是不是也能机械化？

我个人的理解是：所谓机械化，无非是刻板化和规格化。机械化的动作，由于简单刻板，因而可以让机器来实现，又由于往往需要反复千百万次，超出了人力的可能，因而又不能不让机器来实现。因之，机械化为机器化进而自动化铺平道路，是它们必不可少的前奏。就这一意义来说，数学中的某些脑力劳动与体力劳动颇有共同之点，它们也就同样可以机械化。事实上，我们在中小学的课程里，就已学习了不少机械化的数学内容，接受了不少机械化的训练，只是人们并不自觉而已。

举例来说，在小学里用纸笔进行的加减乘除四则运算，就完全是机械化的，正因为如此，才有可能在十七世纪时巴斯喀利用齿轮转动制造成加法机器，稍后莱布尼茨又把它改进成乘法机器。而到现代，四则运算已可在电子计算机上实现。如果没有小学里那种已经成为机械化的算法，这些都将是不可能的。

如果检查一下数学两千年来的发展史，可以看到数学多次重大的跃进都无不与数学的机械化有关。例如算术有许多四则难题，每题求解都需要巧思，但出现了代数以后，许多问题的

---

* 本文摘自《自然辩证法通讯》，1978 年。

入手与解答成为机械性的，变得轻而易举，中学生也很容易学会了。又如欧几里得几何的定理证明，添线往往是一种很高超的艺术，而与机械化毫无共同之处，但出现了解析几何，证明定理就有些机械化而容易入手。再如求极大极小问题，稍难一些就需特殊技巧或无所措手，出现了微积分往往就变为机械地算一下导数就行了。虽然这些都还算不上真正的机械化或半机械化，但总是提高了机械化的程度，在机械化的道路上迈进了一大步，在历史上也都成为数学进展的划时代标志。

在长期的实践中随着人们认识的演化，人们逐渐意识到数学机械化的优越性。17 世纪的莱布尼茨，不仅对数学计算制造了乘法机器，还对数学中的逻辑推理提出了机械化证明定理的设想。到了本世纪，由于数理逻辑的出现而使人们的认识更为明确，又由于电子计算机的出现，因而原来的设想已经进入到如何具体实现的阶段，使数学的发展面临一个前所未有的新时代。

敬爱的周总理早在 1956 年时就已指出，电子计算机已可开始有条件地代替一部分特定的脑力劳动，就象其他机器代替体力劳动一样。我国在体力劳动机械化的革命中曾经落在许多现已成为发达国家的后面，在当前这场脑力劳动机械化的革命中，我国决不能重蹈覆辙，而应努力争取走在世界各国的前面。我国在古代是一个数学大国，它的数学的特点之一是它基本上是一种机械化的数学。缅怀过去，瞻望未来，数学理应充当当前这场新的革命中的一名先锋战士，使我国早日成为一个实现了机械化的现代数学大国。

数学中的公理化与机械化思想*

在全国人民向四个现代化进军的今天，究竟什么是数学的现代化，怎样才能实现数学的现代化，是每个数学工作者不能不关心的问题。为了摸索这一问题的答案，了解一下古今数学思想的发展历史，是不为无益的。

我个人认为：贯穿在整个数学发展历史过程中有两个中心思想，一是公理化思想，另一是机械化思想。前者似已不必多费笔墨，欧几里得几何原本是这方面历史上的代表作，也是公理化思想的滥觞，在现代数学尤其是纯粹数学中占据着统治地位。后者则似有必要稍加说明。

试以中学数学为例。中学的数学课本是一个奇妙的混合物：公理化与机械化的方法内容杂然并陈。欧几里得式的平面几何在整个课程中占据了公理化的一个角落；而代数部分则有丰富的机械化成份。解线性联立方程组所用各种消去法就是典型的机械化方法。30 年代中学数学教本广泛论述的 Sturm 定理无非是代数方程求实根数目的一个机械化方法。至于解析几何，则把几何与代数结合在一起，两种成份都有，但既非此也非彼，而且结合得并不很好，前后并不连贯。例如平面几何费大量篇幅、时间于三角形的全等，但对于解析几何最简单最基本的中点或分点公式，要用到关于梯形中位线或平行分比的定理，一般在中学课本中却很少提及，至少是不够突出，而且由

---

* 本文摘自《课程、教材、教法》，1981 年第 1 期。

于点的位置牵涉到线段的正负，与平面几何部分更显得脱节。如果拘泥于公理化所强调的推理严密性的要求，则一般书中对这一段的论述大都是不合格的。这类有可挑剔之处在中学数学课本中屡见不鲜。

公理化的思想导源于古希腊，机械化的思想则贯穿于整个中国的古代数学。秦汉时代就已成书的《九章算术》，是具有这一思想的代表作，线性联立方程组的解法及有关正负数概念与移项法则，最早就已见于此书。《九章算术》两千多年来一直影响与指导着中国数学的发展，为数学做出了巨大贡献（虽然直到现在还没有被一般人所认识），与欧几里得几何原本东西辉映，各呈特色。中学数学实际上正反映了历史上这两种数学思想与方法的汇合，只是这种汇合还不够合拍，并没有做到浑然一体，有机地溶合在一起。

机械化的思想与方法并不仅见于中小学的数学课本，实际上在现代纯粹数学的研究中也一直发挥着它的作用与影响，举例来说，本世纪的数学巨子法国的 E. Cartan，在他关于微分方程、微分几何与李群的浩瀚著作中，经常显出机械化思维的特色，他的儿子 H. Cartan，是当代数学名家，其最重要的工作之一是关于拓扑学中 $K(\pi, n)$ 同调群的计算，可以看作是运用机械化方法取得成功的一个典范。即使是以公理化著称的数学大师 Hilbert，在他公理化方法的代表作《几何基础》中，全书主要部分的最后一条定理在实质上就是为某一类交点定理提供了一个机械化的证明方法。在他倡导之下蓬勃发展起来的数理逻辑，其作用之一就是在于摸索一条机械化证明数学定理的道路。当前许多数理逻辑学家都重视或从事机器证明的工作，这决不是偶然的。有着两千多年历史成为机械化代表作的线性联立方程组解法，一直是中学数学所不可缺少的组成部分，也是数学实践中最常用到的部分。历史较晚的另一机械化

代表作 Sturm 定理，虽然由于公理化运动的影响以及某些数学学派的轻视，从 30 年代以来逐渐从数学著作中消失，但晚近却又卷土重来，重新出现于一些数学名家的代数学教本之中。随着计算机的出现与其使用日见普及，机械化的方法与作用已日益为数学家所认识。不少在纯粹数学上卓有成就的数学家，正日益重视计算机与机械化方法在纯粹数学中的使用并为之鼓吹。机械化思想的历史功绩是不可磨灭的，它未来的生命力更是无比旺盛的。

公理化与机械化的思想与方法，都曾对数学的历史发展做出了巨大的贡献，今后也仍将继续做出巨大的贡献。我们既不能厚此薄彼，也不能重彼而轻此。为了实现数学的现代化，我们必须吸收渊源于西方的公理化方法的长处，也应珍视我国古代的遗产，从有着历史渊源的机械化方法中吸取力量。这两种方法的融合，或许能为数学的未来发展提供一些新的摸索途径。

以上略陈管见，以作争鸣之用。

几何定理机器证明*

一、几何定理证明——从 Euclid, Descartes 到 Hilbert

定理证明是现代纯粹数学的主要活动形式，导源于古希腊，Euclid 的《几何原本》是其代表著作。即使是平面欧氏几何，就有着无数美妙的定理，以下略举数例，以资说明。

例 1 Feuerbach 定理（1822）。三角形的九点圆与其内切圆 O 以及傍切圆都相切。

例 2 Morley 定理（1900）。三角形三个角的相邻三分角线相交成一个等边三角形。

例 3 Thebault 推测（1938）。设三角形 ABC，L 是过顶点 A 的一直线，则在与底边 BC，直线 L，以及三角形外接圆相切的诸圆中，有两圆其中心联线经过三角形的内心或傍心。

按照通常的 Euclid 模式，几何定理的证明，是依据公理系统，按逻辑规则演绎地进行的。每一个别的定理，其证明只能适合于这一个别的定理。证明没有通用的思想指导，需要有高度的聪明技巧与迂回曲折的推理才能获得，往往表面看来极为平凡的定理，其证明却极其困难，下面是一个著名的例子，所述定理在 19 世纪时曾一度难倒了许多著名的几何学家。

例 4 分角线定理。有两个内分角线相等的三角形必等

---

* 本文摘自《自然科学进展》，1992 年第 1 期。

腰。

顺便指出，以上这些定理在叙述上都有不够完全与模棱两可之处必须加以澄清。这在以后将再加讨论。

在1637年出现了Descartes的《几何学》，1899年又出现了Hilbert的《几何基础》。这两部经典著作为几何定理的证明提供了不同于Euclid模式的可能性。Descartes开创了可用计算来证明几何定理的局面，而Hilbert则指出了几何定理可以不是逐一证明，而是一类定理可以用统一的同一方法一起证明。这一统一的方法，在引入适当的坐标后，也可仅仅用计算来完成，而计算的过程，则可依事先确定的步骤，即今日所称的算法，按步进行以迄于成。应该指出的是，并没有先验的理由来保证这样的算法必然存在，算法的存在与否，其本身也需加以证明。在这样的算法确实存在时，我们将称这类定理的证明是可以机械化的，或简单地说这一类（定理）是可以机械化的。相应的算法则称为这类定理的一个机械化证明。至于算法存在这一事实，则称为一个机械化定理。

Hilbert的思想，已发展成一个新的数学领域：数理逻辑。无数数理逻辑学家，曾致力于各种类型定理的机械化证明。其中最值得称道的成就之一，是Tarski所发现的一个定理，依据我们所用的辞汇，这一定理可表述如下：

Tarski机械化定理（1950） 通常的Euclid几何是可以机械化的。

随着计算机的出现，自然引起如何将一已由某一机械化定理已保证存在的算法，转化为程序使之可在计算机上实现的问题，这方面的一位先驱者是王浩，他在计算机上只用了数秒钟的时间，将Whitchead－Russell《数学原理》一书中的350个以上的命题一举证明。至于Tarski的机械化算法，则由于过于复杂，尽管后来有各种改进与简化，迄今仍未听说有何值得

称道的定理曾获证明。

二、几何定理机器证明的一个有效算法

在 1976 至 1977 年之交，作者发现了对某一类非平凡几何定理的机械化证明方法。为了说明这一方法，先考虑一个特例，例如例 1 中的 Feuerbach 定理（参阅文献 [1]，第 198 页）。

设三角形 ABC 与任一内心或傍心 I，取坐标，使

$I=(0, x_1)$，$A=(x_2, 0)$，$B=(x_3, 0)$，$C=(x_4, x_5)$。

又取九点圆心为 $N=(x_6, x_7)$，而半径为 x_8，记 N 与 I 间的距离为 x_9。容易将几何条件转化为一组方程 $P_i=0$，$i=1, \cdots, 6$，多项式 P_i 见下，记 PS 为由 P_i 所组成的多项式组，则这些方程可简写为 $PS=0$。

$$
\begin{aligned}
P_1 &= (x_1^2+x_2{}^*x_3){}^*x_4-x_1^{2*}(x_2+x_3),\\
P_2 &= (x_1^2+x_2{}^*x_3){}^*x_5-2{}^*x_1{}^*x_2{}^*x_3,\\
P_3 &= 4{}^*x_6-(x_2+x_3+2{}^*x_4), \qquad (1)\\
P_4 &= 4{}^*x_5{}^*x_7-x_5^2-(x_3-x_4){}^*(4{}^*x_6-2{}^*x_2\\
&\quad -x_3-x_4),\\
P_5 &= 4{}^*x_8^2-(2{}^*x_6-x_2-x_3)^2-4{}^*x_7^2,\\
P_6 &= x_9^2-x_6^2-(x_7-x_1)^2。
\end{aligned}
$$

不论 I 是内心或是某一傍心，也不论定理要求证明的相切是内切还是外切，定理的结论总可表达为 $G=0$ 的形式，这里

$G=(x_9+x_8+x_1){}^*(x_9+x_8-x_1){}^*(x_9-x_8+x_1){}^*(x_9-x_8-x_1)$。

注意诸多项式 P_i 逐一引入新变量 $x_4, \cdots, x_9$。这些新变量最高次项的系数将记为 $I_1, \cdots, I_6$。例如 $I_1=x_1^2+x_2{}^*x_3$。

显然如果我们能证明结论方程 $G=0$ 是假设方程组 $P_i=0$ 的后果，则定理即被证明成立。为此使 G 对 P_6，…，P_1 依次取余式以尽可能消除 x_9，…，x_4。最后所得余式将记为

$$R=\mathrm{Remdr}\ (G/PS)。 \tag{2}$$

于是有一如下形式的公式：

$$IP^* G=\mathrm{Sum}_i{}^* \Omega P+R, \tag{3}$$

其中 IP 是诸 I_i 的一个幂积，Ω_i 是商多项式，而 Sum 系对 i 从 1 至 6 求和，由直接计算得

$$R=0。 \tag{4}$$

由（3）与（4）式可见，只需 $IP\neq 0$ 即无一系数 I_i 为零时可从 $P_i=0$ 推得 $G=0$，因而在以下条件

$$I_i\neq 0\ (对所有\ i) \tag{5}$$

之下，Feuerbach 定理已被证明成立。注意条件 $I_1=0$ 意指 AC 与 BC 两直线平行，而 $I_4=4^* x_5=0$ 指顶点 C 将落在 AB 边上而使三角形退化为一直线，为此我们称 $I_i\neq 0$ 为非退化条件。事实上，非退化条件往往隐含在要求证明的定理假设中而不明白说出，乃是一个普遍的情况。我们所用方法，可以自动地探察出这种隐含而未明白表达的非退化条件。

上述 Feuerbach 定理的证明方式具有普遍性质。为对所用方法作一普遍陈述，试设一固定的特征为零的域 K 与一组变量 x_1，…，x_n。在环 $R\in K\ [x_1,\ \cdots,\ x_n]$ 中的任一非零多项式 P 中，所出现的 x 的最大下标将称为 P 的类。在 P 中不出现任一 x 即 P 是一非零常数时，P 的类将置为零，对于一类 $c>0$ 的多项式 P，恒可书写成下面的典范形式：

$$P=I^* x_c^d+x_c\ 的低次项,$$

其中诸 x_c 幂的系数本身也是 $K\ [x_1,\ \cdots,\ x_{c-1}]$ 中的多项式，这时指数 d 将称为 P 的次数而领导项系数 I 将称为 P 的初式，

不仅是通常的欧氏几何，在各种不同的几何中，大量的定理在选择适当的坐标后，可以容易地表现为下面的形式。

定理 T 的假设可表现为一组方程 $P_i=0$，这里 P_i 是类为 $c_i>0$ 的假设多项式，其典范形式可写为

$$P_i = I_i^* x_{ci}^{di} + x_{ci}\text{的低次项},\quad i=1,\ \cdots,\ r。\tag{6}$$

这里假定

$$0<c_1<c_2<\cdots<c_r。\tag{7}$$

通常还假设以下条件（C）也成立：

（C）对于任两 $j<i$ 的 i，j，P_i 的初式 I_i 中 x_j 的次数小于 d_j。

这样的多项式组 $PS=\{P_i\}$ 将称为一个升列，并称其型是 0 或 1，视其是否满足（7）式或同时满足（7）式与（C）而定。这一概念的重要性，从与（1）式中多项式组的比较可见一二。

今设定理 T 的终结也可表示成 $G=0$ 的形式，这里 G 是 R 中的终结多项式。今如（2）式将 G 对 PS 中诸多项式依反向次序 P_r，$\cdots$，P_1 依次求余，以作余式 $R=\mathrm{Remdr}(G/ps)$。于是可得一（3）式那样的余式公式，如前显然可知在 $R=0$ 时，只需（5）式成立，即可从 $P_i=0$ 得出 $G=0$。条件 $R=0$ 不仅是使在（5）式条件下得出 $G=0$ 是充分的，而且在 PS 是型 1 升列且不可约的情况下，也是必要的。这里 PS 不可约的意义是，它具有型 1 且方程组 $PS=0$ 的方程。依次对由这些方程定义的 K 的扩张域中不可约，这些结果可表成如下形式的定理。

定理 1　对于一个以多项式方程组 $PS=0$ 为假设，多项式方程 $G=0$ 为终结的定理 T，这里 PS 是一升列，可依（2）式作余式 R，于是 $R=0$ 是定理 T 在由（5）式给出的非退化条件下成立的充分条件，在 PS 是型 1 的不可约升列时，也是

必要条件。

上述证明几何定理的原理，表面看来极其平凡，但已证实极为有效，例如，在 SUN3/140 工作站上用所附 REDUCE 软件包进行计算时，上述 Feuberbach 定理的证明只需 2.7s。作为另一实例，不妨考虑例 2 中的 Morley 定理，试先注意对于任两依一定先后次序的相交直线所成的角，应有 3 个主三分角线，以及相伴的三个次三分角线。对于一个三角形的三个角，它们的相邻三分角线将交成 27 个三角形。用以上方法可以证明其中恰有 18 个三角形是等边三角形。证明的计算是在一台 HP9835A 的台式计算机上完成的，它的内存只有 256K（参阅文献 [2]）。

在美国与本国，各已有几百条不平凡的定理在相对较小的计算机上应用上述方法获得证明，可参阅吴文俊、王东明、高小山与周咸青等的有关论著。最大规模的实验，则由周咸青做出[3]。

我们的方法虽局限于用多项式表达的定理，但即使是三角恒等式，也可用这一方法来证明，只需事先把三角函数化为有理函数即可。例如在作者最早的有关论文[4]中，即已举例如下。

例 5 对于三角形 $A_1A_2A_3$ 的 3 个角，有

$$\sin(2^* A_1)+\sin(2^* A_2)+\sin(2^* A_3)$$
$$=4^* \sin A_1{}^* 1\sin A_2^* \sin A_3。$$
$$\cos(2^* A_1)+\cos(2^* A_2)+\cos(2^* A_3)$$
$$=-4^* \cos A_1^* \cos A_2^* \cos A_3。$$

既然有了一个证明几何定理的有力方法，就可致力于几何新定理的发明发现。以下两定理都为作者所发现，很可能是前人所未知的（参阅文献 [1，2]）。

例 6 Pascal 锥线定理对于在同一圆锥曲线上的 6 个点 A_1，…，A_6，称任取其中 4 点所成两弦的交点为 Pascal 点。这样的点有 45 个，对文字 1，…，6 的任一置换 s，将引出这些 Pascal 点之间的一个置换，仍记之为 s，于是对于 A_1A_3 与 A_2A_5 的 Pascal 交点 P，诸点 P，sP，…，s^5P 将在同一圆锥曲线上。

例 7 设一平面在一固定平面上移动而占有4个不同位置，设对第 i 个与第 j 个位置间的运动是一依所谓极点 P_{ij} 为中心的一个旋转，于是 $P_{12}P_{34}$，$P_{13}P_{24}$，$P_{14}P_{23}$ 的 3 个中点必在同一直线上。

当例 7 中 6 个极点 P_{ij} 所构成的图象退化为一完全四边形时，相应的定理即成为 Gauss 所发现而为 Gauss 自己所赞赏的一个定理（参阅文献 [5，6]）。

最后试注意在定理 1 中所出现的非退化条件（5），在某种意义下是不可少的。事实上，如果不附加这种非退化限制条件，则定理可以成为无意义甚至根本不成立，即使通常认为是成立的，对于这一事实所引起的严重后果，可以参阅作者在文献 [7] 中的某些议论。同样，由于理想难以处理非等式，因而用理想的概念与理论来处理几何定理的机械化证明，看来是不适当的。但在另一方面，非退化条件这一概念在某种意义下又不是不可或缺的，其理由是：如果定理的假设多项式组 $PS=\{P_i\}$ 是一型 1 的不可约升列，则方程组 $PS=0$ 将定义一不可约代数簇 V，这一代数簇 V 除包含所有使 $PS=0$ 但无一初式 $I=0$ 的点之外，还可能包含某些使某一 $I_1=0$ 的满足 $PS=0$ 的点，以 R 表（2）式所定的余式，其中 G 是定理 T 的终结多项式，则 $R=0$ 将是定理对任一相应于 V 上任一点的几何图象都能成立的既必要又充分的条件，而 V 既包含那种满足 $PS=0$ 而使所有 $I_i\neq0$ 的点，又包含某些满足 $PS=0$ 但

也使某一 $I_i=0$ 的点，由此可知，非退化条件的引入，完全可以避免。再者，如果认为需要，可以通过计算定出 V 的周炜良形式与周炜良坐标，并由此定出与 V 相应的理想的一组基。这一组基中的多项式，通常与 PS 中的多项式并不相同，这样一组基的存在与确定，虽然有丰富的理论意义，但无补于几何定理的证明，也对定理 T 究竟在何种几何图象上成立并不能提供清晰的了解。正因为如此，对于定理的机械化证明，我们宁愿在定理 1 中引入非退化条件，而满足于那种较不确切的叙述方式。

三、几何定理机械化证明的一般原理

没有什么先验的理由，可以保证一个几何定理的假设与终结都必然用多项式方程来表达。即使都能用多项式方程来表达，也没有什么先验的理由，可以保证假设多项式组必然以升列的形式出现。对于前者，作者曾经证明，对于多数由公理系统出发的重要的几何，其有意义的几何概念与几何关系，都可用多项式、式项式方程，多项式非等式或在牵涉到次序关系时的多项式不等式来表示。所出现的多项式，其系数都在一基本域内，而这一基本域，是由从公理系统出发的几何内在地确定的。这不仅对从属于通常欧氏几何的一系列几何来说是如此，对向来大量使用三角函数、双曲函数的非欧几何来说也是如此(详细可参阅文献［1］)。

基于上述理由，我们在此后将设所考虑不牵涉次序关系的几何定理 T，其假设与终结将作 $HYP=0$ 与 $CONC=0$ 的形式，这里 HPY 与 $CONC$ 各为环 $R=K\left[x_1, \cdots, x_n\right]$ 中的多项式组与多项式，而 K 是内在地联系于几何的域，我们将假设其特征为零，这时定理 T 将记作 $T=\{HYP, CONC\}$，一般说来 HYP 是由若干杂乱无章的多项式所组成，依据前节，

我们的首要任务乃是将 HYP 转化成一个井然有序的多项式组，例如升列形式，这样就可应用前节所说的方法。

前面已经指出，用理想的概念与方法来处理 MTP 并不见得合适。另一方面 HYP 的一个零点或 $HYP=0$ 的一个解，即使是在 K 的一个扩域中，也即是满足定理 T 的假设的一个几何图象，最多这一图象存在于扩充后的空间而已。所谓要证明定理 T，无非是说要确定这些图象全体中的这样一部分，使在这一部分的图象上，都有 $CONC=0$。为此我们将避开使用以 HYP 为基的理想这种概念，而引入 R 中多项式组 PS 的零点集的概念，记作 Zero（PS），其中包括所有 K 的有牵涉的扩域中的零点，若另有一 R 中的多项式 G，则我们将置

$$\mathrm{Zero}(PS/G)=\mathrm{Zero}(PS)-\mathrm{Zero}(G)。$$

特别有以下诸式：

$$\mathrm{Zero}(PS/0)=\text{空集},$$

$$\mathrm{Zero}(PS/1)=\mathrm{Zero}(PS),$$

$$\mathrm{Zero}(P_1^* P_2)=\mathrm{Zero}(P_1)+\mathrm{Zero}(P_2), P_1, P_2\in R。\tag{8}$$

我们将依赖 Rilt 关于微分方程两书[8,9]中的概念与方法，来完成将任意多项式组 PS 引至一其零点集与 PS 紧密相关的升列的工作，为此试先将升列的概念略加扩大，以包括只含有一个非零常数的多项式组的情形。这样的升列将称为是平凡的。显然一个平凡升列的零点集是空集。

下面关于特征列的概念，将在整个理论中扮演中心的角色。

基本定义　一个升列 CS 将称为多项式组 PS 的特征列，如果下面（9）和（10）两式成立：

$$\mathrm{Zero}(PS)\subset\mathrm{Zero}(CS),\tag{9}$$

$$\mathrm{Remdr}(P/CS)=0, \text{对所有 } P\in PS。\tag{10}$$

从这一定义立即可推得下面的

基本引理 设 CS 是多项式组 PS 的一个特征列。命 I_i 为 CS 中诸多项式的初式，而 J 为这些初式的乘积，则有

$$\mathrm{Zero}(PS/J)\subset\mathrm{Zero}(PS)\subset\mathrm{Zero}(CS), \tag{11}$$

$$\mathrm{Zero}(PS)=\mathrm{Zero}(CS/J)+\mathrm{Sum}_i\mathrm{Zero}(PS+I_i)。 \tag{12}$$

今在 R 中所有非零多项式间，先依其类再依其次数，引入一部分次序。其次，在所有平凡或非平凡升列之间，引进一部分次序，引进的方式将从略。对于一个多项式组，它所能包含的最低序次的升列将称为它的基列。运用这些概念，我们构造特征列的主要程序，即可依据以下格式（S）来完成：

$$\begin{array}{ll} PS=PS_0 & PS_1\cdots PS_m, \\ BS_0 & BS_1\cdots BS_M=CS, \\ RS_0 & RS_1\cdots RS_m=\text{空集}。\end{array} \tag{S}$$

在格式（S）中，每一 BS_1 是 PS_i 的一个基列，每一 RS_i 是 PS_i-BS_i 中的多项式对升列 BS_j 的所有可能有的非零余式所成的集合，又对每一 i，

$$PS_i=PS_{i-1}+RS_{i-1}。 \tag{13}$$

容易证明，这一过程在有限步后设为第 m 阶段时必然中止，这时的 $RS_m=$ 空集，而 BS_m 即为具有性质（11）与（12）式的 PS 的一个特征列。

为了减轻计算的劳动量，也可将上述过程略加修改，使格式（S）中 PS_i 不依（13）式而依下面（14）式来确定：

$$PS_i=BS_{i-1}+RS_{i-1}。 \tag{14}$$

如果对 PS 中任一多项式 P，有余式 Remdr（P/BS_m）$=0$，则最后所得的 BS_m 仍为一所求的特征列。如其不然，可将所有非零余式添入 BS_m 以得一新多项式组 PS'。然后将 PS 易为 PS' 仍依格式（3）进行。如此反复进行，最后可得一所需的特征列（参阅例如文献［10］）。

今将以上结果表达如下：

分解定理Ⅰ 有一算法使对任一多项式组 PS 在有限步内确定 PS 的一个特征列 CS，满足（11）与（12）式。

对（11）式中 Sum 下的每一零点集都依次做同样处理，如此反复进行，最后即可得到下述定理：

分解定理Ⅱ 有一算法使对任一多项式组 PS 在有限步后获得有限多个升列 AS_i，使

$$\text{Zero}(PS)=\text{Sum}_i\text{Zero}(AS_i/J_i),\tag{15}$$

其中 J_i 是升列 AS_i 中诸多项式的初式的乘积。

在以上进程中若出现可分解因子的多项式，即可依（8）式将零点集分解并如前进行。如此反复进行，最后即可获得一形式仍如（15）式的分解，但其中每一升列 AS_i 都是不可约的，今改记为 IRR_i。对每一这样的不可约升列 IRR_i，试作 R 中对 IRR_i 有零余式的所有多项式所成的理想。这一理想所成的零点集是通常意义下以 K 为基本域的一个不可约代数簇，记之为 $\text{Var}[IRR_i]$。注意这里用的是方括号。这是因为，为了区别于以 IRR_i 为基的理想所定的代数簇，后者不妨记之为 $\text{Var}(IRR_i)$。这两个代数簇并不相同。应用这些定义与符号，即可将上述分解定理深化如下。

分解定理Ⅲ 有一算法使对任一多项式组 PS 在有限步内确定一组不可约升列 IRR_h，使

$$\text{Zero}(PS)=\text{Sum}_k\text{Var}[IRR_k]。\tag{16}$$

而且，仅仅通过计算，即可将以上 Sum 中不必要的多余成分删去，而使 SUM 中每一代数簇都不含于另一代数簇之内。

注意（16）式给出了代数簇 Zero（PS）的一个不可缩的不可约分解，显然除次序外这个分解是唯一的。

今将分解定理中的 PS 取为几何定理 $T=\{HYP, CONC\}$ 中的 HYP，则对每一分解定理，有一机证原理

（MTP 原理）与之相应，而第一个原理 I，可直接导致第二节中的定理 1。

机证原理 I 对任一定理 T＝｛HYP，$CONC$｝，命 CS 是 HYP 的一个特征列。又命 R = Remdr（$CONC/CS$）是 $CONC$ 对 CS 的余式，则 $R=0$ 是使定理 T 在非退化条件（5）下普适成立的充分条件，而在 CS 是不可约升列时，也是这一事实的必要条件。

机证原理 Ⅱ 对一定理 T＝｛HYP，$CONC$｝，命 AS_1 是一组由 HYP 导出的升列，使

$$\text{Zero}\,(HYP) = \text{Sum}_i \text{Zero}\,(AS_i/J_i), \tag{17}$$

其中 J_i 是升列 AS_i 中诸多项式的初式的乘积，则在

$$\text{Remdr}\,(CONC/AS_i) = 0 \tag{18}$$

时，定理 T 将在 Zero（HYP）中的 Zero（AS_i/J_i）部分上成立，即有

$$\text{Zero}\,(AS_i/J_i) \subset \text{Zero}\,(CONC)。 \tag{19}$$

机证原理 Ⅲ 对一定理 T＝｛HYP，$CONC$｝，命 IRR_k 是一组由 HYP 导出的不可约升列，使

$$\text{Zero}\,(HYP) = \text{Sum Var}\,[IRR_k], \tag{20}$$

则定理 T 在不可约的成分 Var［IRR_k］上成立的充要条件是

$$\text{Remdr}\,(CONC/IRR_k) = 0。 \tag{21}$$

注意在原理 Ⅰ 与 Ⅱ 中，都有非退化条件，各以 $J\neq0$ 与 $J_i\neq0$ 的形式出现，但在原理 Ⅲ 中，已不再有任何非退化条件。再者，原理 Ⅲ 对定理的成立与否还提供了最完全的信息，例如，试考察三角形 3 个角的分角线交于一点这一定理，在适当坐标下，假设多项式组的零点集将分解为 8 个不可约成分。这与三角形每一角两个分角线，因而 3 个角的 3 个分角线可有 8 种不同的情形相当。应用判据（21）式可知在 8 个不可约成分

上，其中 4 个定理成立而其它 4 个则否。这与三角形有 4 个内心与傍心的几何事实相当。另外一面，获得（20）式那样完全的分解由于需要在代数域上分解因子，往往极为复杂困难，所以我们往往宁愿应用较不完全的原理 I 与 II。事实上，大多数迄今为止所证明或曾经发现的不平凡的几何定理，只应用了最简单的原理 I。此外，作者为了排除在初等几何定理中经常出现的各种模棱两可，曾经引入有向直线与有向圆的概念。应用了这些概念，假设多项式往往可以分解因子，这相当于将基本域 K 适当扩充，使原来只是相对不可约的代数簇分解成为绝对不可约的代数簇，关于有关概念可参阅文献［11］。作为说明，试考虑例 3 中的 Thebanlt 猜测。这一猜测系 Thebanlt 在 1938 年时提出，但直到 1982 年才由 Taylor 用极其繁复的经典方法证明。这一猜测中含有许多模棱两可之处，因为与两直线及一圆相切的圆有 4 个，与三直线相切的圆也有 4 个，因而猜测中的圆应如何选择，才可使其中心位于一直线上，在猜测的叙述中并不明确，在这方面周咸青首次在机器上较为明确也更为一般地证明了上述定理。周所用的机器是 Symbolics，所用机证的 CPU 时间为 44h，而 maxt = 674，927，这里 maxt 指机证过程中所出现多项式的最大项数。它提供了一个比 CPU 更为客观的证明复杂度的指标。在周证后不久，作者即依据原理 III 以及有向直线与有向圆的概念，将这一 Thebault - Taylor 周定理重证一遍，所用机器是只有 1M 内存的 Dual 机，实际运行时间约为 6h，而 max$t$ = 4885。同样的方法也曾用之于 Feuerback 定理及若干其它定理，参阅文献［12］可知其详。估计这一方法应可成功地应用以探讨若干精致而复杂的图象，例如由与三圆相切的 8 个圆所组成 Hart 图象等。最近，杨路、张景中及其追随者们在这类所谓可约性的几何定理证明方面做出了有意义的贡献。

最后，试注意对于一升列 AS，零点集 Zero（AS）可以认为已完全确定或方程组 $AS=0$ 可以认为已经解出，因而分解定理Ⅰ—Ⅲ可以作为解任意多项式方程组 $PS=0$ 的基础（参阅文献 [13, 14]），由于科学与技术中的大量问题往往归结为多项式方程组的求解，这些定理的应用范围之广阔无垠，自不待论，但本文仅局限于这些定理在几何定理机器证明上的应用以及与之紧密相关的某些问题。

四、定理的自动发现与对微分几何的推广

上面所描述的方法，不仅可用以证明已知的几何定理，还可用以自动发现新定理或未知关系，所依据的原理是很简单的。在第三节主要算法的格式（S）中，所得诸 PS_1，BS_1 或 RS_1，尤其是最后所得 CS 中的多项式都属于以所给多项式组 PS 为基的理想，因而对于这些多项式 P 都有

Zero（PS）$\subset$Zero（P），$P\in PS$，BS_1，RS_1 或 CS。

今设 $PS=HYP$ 是一多项式组，其零点相应于满足某些几何条件的几何图象。假设从 HYP 确定其特征列 CS 的过程中，出现一多项式 P，而方程 $P=0$ 具有某种几何意义，则置 $P=CONC$ 时，即有一形如 $T=\{HYP,\ CONC\}$ 的几何定理，而这一定理的成立，并不受任何非退化条件的限制。

上面在算法过程中出现具有几何意义的多项式方程 $P=0$ 一事，带有某种偶然性，但我们可以利用这一点，用以下方法来有意识地自动导致有几何意义的结果。首先注意假设多项式组 HYP 的特征列 CS，是一升列，不妨假设为非平凡的，于是 CS 将由若干多项式所组成，设为 C_1，C_2，…，C_r，它们逐一引入新变量 x_{C_1}，…，x_{C_r}，这里 c_i 是 C_i 的类而 $0<c_1<\cdots<c_r$。今设可以事先肯定在某些几何量之间必具有某些关

系，或希望求得某些几何量之间的可能有的关系，这时可有意识地将诸变量 x_i 如此排列，使需要求得某种关系的那些几何量成为下标最小的那些 x，其它的几何量则排列为下标较大的 x。于是通常所获得特征列 CS 中的第一个多项式 C_1 即将给出所求的关系 $C_1=0$。

我们曾应用上述极为简单的原理于新定理与未知关系的自动发现。最简单的应用是三角形面积与三边之间 Heron 秦九韶关系的自动发现，这里对于可能具有的关系的形式不作任何假定。同样我们也自动发现了较少为人所知的四面体体积与六棱之间的关系式。我们甚至自动发现了双曲平面中三角形面积与其三边之间的关系。注意在第二节中就已指出，在这一问题中出现的三角函数与双曲函数等超越函数，对应用我们的方法并无妨碍。所得结果见下。关于这些以及其它实例可参阅文献〔15〕。

例 8 双曲面积定理。边长为 s_1，s_2，s_3 的双曲三角形的面积 A 由下式给出，其中 $x_1=\cosh s_i$，而 $y=\cos A$：

$$(x_1+1)^*(x_2+1)^*(x_3+1)^*(y-1)$$
$$=-2^*x_1{}^*x_2{}^*x_3+x_1{}^2+x_2{}^2+x_3{}^2-1。$$

我们关于 MTP 的整个理论与方法，并不限于初等几何，而已推广到微分几何。事实上，微分情形与初等的情形几乎是平行的。自然，这时通常的域需改为微分域，通常的多项式需改为微分多项式，等等。同样，在初等情形下的变量 x_i 需改为对某一组独立变量的函数与其导函数，于是，相应于第三节中的分解定理与 MTP 原理，我们有完全类似的定理与原理。其细节自然要复杂得多，尤其是在独立变量不只一个时是如此，因为那时必须考虑到完全可积的条件。整个理论依赖于过去 Riquier，Janet，E。Cartan，与 Thomas 等人的工作，我们将不作介绍，请读者参阅本文文末所列的有关文献，特别是作

者的文献〔16〕。

显然正像初等几何那样，可以在微分几何中证明甚至发现大量的有趣的定理。但由于计算极其复杂，因而迄今为止，我们的工作基本上局限于一个独立变量即常微的情形或几何中相当于曲线论的情形。

下面是一个我们曾经详尽探讨过的情形。详见文献〔17〕。

例9 Bertrand 曲线偶定理。在通常的三维空间中，设有两成一一对应的曲线 C，C' 在对应点的主法线重合，则有以下诸性质：

(a) 对应点间的距离是一常数；

(b) 对应点切线的交角是一常数；

(c) (Bertrand) C（或 C'）的曲率与挠率有一线性关系；

(d) (Scheii) 对应点的挠率的乘积是一常数；

(e) (Mannheim) 对应点及其联线上曲率中心的交比是一常数。

这些定理不仅可以作为已知定理加以证明，并可在对它们一无所知的假定下应用本节所说明的技巧自动地发现出来。下面试就本例做一简单的描述。

在曲线 C 的每一点 X 处取一活动标架 $(X,\ e_1,\ e_2,\ e_3)$，其中 e_i 是 X 处有特殊意义的三条直线上的单位向量，这三条直线是切线，主法线与从法线。命 s 为 C 的弧长参数，而 k，τ 各为 C 的曲率与挠率，则有 Frenet 方程如下：

$$\frac{dX}{ds}=e_1,\quad \frac{de_1}{ds}=ke_2,$$

$$\frac{de_2}{ds}=-k^*e_1+\tau^*e_3,\quad \frac{de_3}{ds}=-\tau^*e_2。$$

在 C' 的对应点 X' 处也可对弧长参数 s' 作活动标架及其 Frenet 方程，其中含有 C' 在 K' 的曲率 k' 与挠率 τ'。此外在相应活动标架之间，又有如下形式的关系：

$$X' = X + a_2^* e_2,$$
$$e'_i = \mathrm{Sum}_j u_{ij}{}^* e_j,$$

其中（u_{ij}）是一正交矩阵，使

$$u_{11} = u_{33}, \quad u_{13} = -u_{31},$$
$$u_{12} = u_{21} = u_{23} = u_{32} = 0,$$
$$u_{11}^2 + u_{13}^2 = 1。$$

置$\frac{ds'}{ds} = R$，于是比较两组 Freuet 方程，易得一组方程 $DP_k = 0$，其中 DP_k 是以下诸函数

$$R, \ a_2, \ u_{ij}, \ k, \ \tau, \ k', \ \tau'$$

及其对 s 导数间的微分多项式。它们所成微分多项式组 $DPS = \{DP_k\}$ 的零点或微分方程组 $DPS = 0$ 的解（在任意基本微分域的有关微分扩张域中），即是符合所设几何关系的某种几何图象。

今设我们的兴趣在于探求两曲线对应点的挠率之间可能有的关系，而关系如何，事先一无所知。这时我们可将前面所列的函数记成 x_i，而使 τ，τ'是其中有最小下标的两个 x，例如 x_1 与 x_2，今如初等情形那样，作出微分多项式组 DPS 的一个微分特征列，可以发现它的第一个微分多项式 C_1 具有以下形式：

$$C_1 \frac{dx_1}{ds}^* x_2 + \frac{dx_2}{ds}^* x_1。$$

显然 $C_1 = 0$ 等价于

$$\tau^* \tau' = 常数。$$

这样，我们就以自动方式发现了（d）中的 Schell 定理，而事先并不知道其形式如何。其它诸定理也可同样处理。

应用同样方法我们曾探讨过这样的成一一对应的曲线偶，其中一曲线在一点的某一有特殊意义的直线与另一曲线在对应

点处的某一有特殊意义的直线重合。关于这种曲线偶的某些已知定理已重新证明或被自动发现。这些尝试并已施之于三维仿射空间中的曲线偶，在近年，周咸青、高小山与李子明对曲线论已用这一方法做了大量研究工作。

上述自动发现的方法还曾施之于 Newton 引力定律与 Kepler 观察定律间的相互关系，作者在此应向美国 Argonne 实验室 Gabriel 教授提出这一问题致谢。为考虑这一问题，试取适当坐标以太阳为原点，将 Kepler 定律转化为以时间为独立变量的微分方程组，并将出现的函数记为 x_i，其中日地距离与加速度定为下标最小的两个 x，在确定微分特征列的算法运行过程中，可发现某微分基列中有一微分多项式，置为零时的方程，即相应于 Newton 的反平方律，其详细情形可参阅文献〔18〕。所得上述结果可表达如下：

例 10 行星运动定律。Newton 的反平方律是 Kepler 观察定律的逻辑后果，它可从后者自动推出，而不必事先对其形式有所了解。

五、含有次序关系或不等式的几何定理机器证明

Tarski 的机械化定理可用于实域中通常几何的机证，这种几何的一个特征是含有次序关系，在代数表达上相当于不等式，前已指出，Tarski 方法的实效不高。Collins 在 1973 年创立了一种代数分解方法（简称 CAD 法），是对 Tarski 方法引人注目的极大改进。近年来，这一方法已被成功地应用于某些实际问题，例如所谓碰撞问题，即确定在不同途径上运动的两物体是否会发生碰撞的问题。这说明方法前景可观，但在另一面，就几何定理机证而言，这一方法的效果如何似乎还有待检

验。

至于我们的方法，基本上局限于不含有次序关系或代数上不含有不等式的那种几何定理的机器证明，因而我们方法的应用范围似乎要比 Tarski 与 Collins 方法陕隘得多。另一面我们方法却有着高效的优点，并可直接推广至微分几何。也应该指出，我们方法的限制并不是说根本上不可能考虑含有次序关系或代数上含有不等式的那种几何定理。事实上，命 R 为实数域，P 是 R〔x_1，…，x_n〕中的一个实系数多项式，则 $P>0$ 或 $P\geqslant 0$ 各与下述事实相当：存在某一新的实变量 y 使 $y^2 * P=1$ 或 $P=y^2$。由此不等式关系可转化为存在性的等式关系。早在文献〔1〕中，作者就已指出如何含有次序关系的几何定理，可以恁借上述转化不等式为等式的技巧，再应用我们的方法来加以证明，书中并举有若干实例。周咸青曾用此法证明过不少定理，高小山还引入某些特殊的技术来证明这一类定理，特别是，周曾与作者合作，彻底解决了例 4 中的分角线定理。容易举例，当三角形的两个相等分角线中至少有一个是外分角线时，三角形不必等腰。只有在两个相等分角线都是内分角线并将这一事实作适当数学表达后再应用我们的方法，才能证明三角形必等腰。证明的复杂性在于在过程中出现的某些多项式因子必须证明是恒正的。试参阅文献〔2〕与通俗小册子[19]。由于迄今并无确定一多项式是否恒正的有效一般方法，因而上述方法还不能普遍应用于含有不等式的所有定理，特别对终结是以不等式表达的定理来说是如此。

1985 年，作者发现了一种证明不等式的方法，它依据下面简单的考虑。

设要证明在某些以等式或不等式表现的限制条件下的某一不等式 $p>0$ 或 $P\geqslant 0$。显然这一问题等价于要证明在上述限制条件下 P 的最小值 >0 或 $\geqslant 0$。必要时引入某些新变量与新

方程，即可将上述问题，除了有时还需确定变量无穷大时的值以外，转化为通常很少几个下述型问题的综合：在某些多项式等式条件例如 $h_1=0$，…，$h_m=0$ 之下，求 P 在一开区域 D 中的局部极小值，于是 P 的最小值即是分成若干个同类问题后所得局部极小值的最小值，于是只需证明这一最小值必 >0 或 $\geqslant 0$ 即可。至于每一个分成的问题，则可用下述方法解决，取 Lagrange 乘子 λ_1，…，λ_m，并作 Lagrange 多项式

$$L=P+\mathrm{Sum}_j\lambda_j^* h_j。$$

命 LAG 为多项式组

$$LAG=\left\{\frac{\partial L}{\partial x_i},\quad h_j\right\}。$$

又在变量 x_i 中任取 m 个而作 h_1，…，h_m 对这 m 个变量的 Jacobi 行列式，记 JAC 为所有这些行列式及 h_j 所成的多项式组，即

$$JAC=\left\{\frac{\partial\ (h_1,\ \cdots,\ h_m)}{\partial\ (x_{i_1},\ \cdots,\ x_{i_m})},\quad h_j\right\}。$$

从初等微积分可知，所有开区域 D 中在条件 $h_j=0$ 下 P 可能取的局部极小值，必然是在 D 中属于 Zero（JAG）或 Zero（JAC）的零点处。这两零点集可由第三节中的分解定理定出。因而所有 D 中可能的极值点与极值，包括局部极小值及其所在的点，可以定出。由此即可定出 P 的最小值来。

我们曾用这一极其简单而初等的原理来处理过各种类型涉及不等式的问题，例如多项式的恒正问题，以下试举在文献〔20〕中的数例，以略见这一原理应用的范围及其有效程度。

例 11　三角不等式。对一三角形 ABC，有

$\sin A+\sin B+\sin C\leqslant 3^*\mathrm{sqr}\ (3)\ /2$,

$\sin A^*\sin B^*\sin C^*\leqslant 3^*\mathrm{sqr}\ (3)\ /8$,

$\cos A+\cos B+\cos C\leqslant 3/2$,

$$\cos A * \cos B * \cos C * \leqslant 1/8,$$

$$\cos A * \cos B + \cos A * \cos C + \cos B * \cos C \leqslant 3/4。$$

再者，上述诸不等式在且仅在三角形等边时才能成为等式。

例 12 Pedoe 不等式。设 ABC，$A'B'C'$ 是在同一平面中的两个三角形，其诸边各为 a，b，c 与 a'，b'，c'，而面积各为 Δ 与 Δ'，则有

$$a'^{2} * (b^2+c^2-a^2) + b'^{2} * (c^2+a^2-b^2) + c'^{2} * (a^2+b^2-c^2) \geqslant 16 * \Delta * \Delta'。$$

再者，等式只能在两三角形相似时成立。

上述方法也曾直接应用于涉及次序无关的几何定理的证明。简单的例子已足以说明这一方法比所有已知的方法都更为有力，但尚需进一步实验，以确知是否像前几节所说对不含次序的定理那样，同样可以证明高难度的定理，作为我们方法的一个挑战，乃是下面由 Kazarinoff 在 50 年代末提出的一个推测（参阅文献〔21〕，第 78 页）。

例 13 Kazarinoff 推测，一个三角形内接于一给定的三角形将其分成 4 个小三角形，则内接小三角形的周界不能同时都小于其它 3 个小三角形的周界。

上述方法还曾应用于各种优化问题，特别是非线性规划问题，一个典型问题，所谓化学平衡问题，已曾做过详尽的处理（参阅文献〔22〕），对于带有微分的优化问题，也可同法处理，但尚有待于进一步实验，其它与本文有关的问题，读者还可参阅文献〔23〕—〔44〕。

参 考 文 献

〔1〕 吴文俊，几何定理机器证明的基本原理（初等几何部分），科学出版社，1984。

〔2〕 吴文俊，系统科学与数学，4（1984），207—235，*Republished in* J.

Automated Reasoning, 2 (1986), 221—252。

〔3〕 *Chou. S. C.*, Mechanical geometry theoremproving, Reidel, 1988.

〔4〕 *Wu Wen-tsun*, Scientia sinica, 21 (1978), 159—172.

〔5〕 *Gauss, K. F.*, Werke, 4 (1882), 385—392.

〔6〕 *Wu Wen-tsun*, J. Svs. Scis & Math. Sci., 2 (1989), 97—109.

〔7〕 吴文俊，数学学报，2 (1982), 125—138。

〔8〕 *Ritt. J, F.*, Dtfferential Equations from the Algebraic Standpoint. *Amer. Math. Soc.*, 1932.

〔9〕 ______, Dtfferential Algebra, Amer. *Math. Soc.*, 1950.

〔10〕 *Wu Wen-tsun*. MM-Res. Preprints, 1989, 3:27—29.

〔11〕 *Hodge. W. V. D. & Pedoe. D.*, Methods of Algebraic Geometry, *Vol. 2*, *Cambridge*, 1952.

〔12〕 吴文俊，中国数学季刊，2 (1987); also in*MM-Res. Preprints*, 1987, 2:18—36。

〔13〕 吴文俊，科学通报，31 (1986), 1—5.

〔14〕 Wu Wen-tsun, *MM-Res. Preprints*. 1987, 1:2—12.

〔15〕 吴文俊，系统科学与数学，6 (1986), 204—216；科学通报., 31 (1986), 436—440。

〔16〕 吴文俊，系统科学与数学，2 (1989), 289—312。

〔17〕 吴文俊，科学通报，32 (1987), 585—588。

〔18〕 Wu Wen-tsun, *MM-Res. Preprints*, 1987, 1:53—61.

〔19〕 吴文俊、吕学礼，分角线相等的三角形，人民教育出版社，1985。

〔20〕 Wu Wen-tsun, *MM-Res. Preprints*, 1987, 2:1—17.

〔21〕 Kazarinoff, N. D., *Geometric Inequalities*, Math. Ass. Amer., 1961.

〔22〕 Wu Wen-tsun, *MM-Rreprints*, 1989, 4:22—39, also Acta Math. Scientia, 4 (1990), 361—374.

〔23〕 Bledsoe, W. W. & Loveland, D. W. (Ed.), *Automated Theorem Proving, after 25 Years*, Amer. Math. Soc., 1984.

〔24〕 Cartan, E., *Les Systemes Differentiels Exterieurs et Leurs Applications Geometriques*. Paris, 1945.

〔25〕 Chou, S. C., Proving elementary geometry theorems using Wu's algorithm, in 〔23〕, 243—286.
〔26〕 ________, Proving and discovering geometry theorems using Wu's method, Ph. D. Thesis, Univ. of Texas at Austin, 1985.
〔27〕 高小山，系统科学与数学，7 (1987), 264—272。
〔28〕 Gao Xiao-shan, *MM-Res. Preprints*, 1987, 2:37—47.
〔29〕 ________, *ibid.*, 1989, No. 3.
〔30〕 Janet, M., Lecons sur les systemens d'equations aux derivees partielles, Gauthier Villars, 1920.
〔31〕 Riquier, C. H., La methode des fonctions majorantes et les systemes d'equations aux derivees partielles, Gauthier Villars, 1928.
〔32〕 Thomas, J. M., *Differential Systems*, *Amer. Math. Soc.*, 1937.
〔33〕 ________, *Annals of Math.*, 30 (1929), 286—310; 35 (1934), 306—311.
〔34〕 Wann, D. M., *Journal of Geometry*, 36 (1989), 173—182.
〔35〕 ________, *MM-Res. Preprints*, 1987, 1:38—52.
〔36〕 ________, *ibid.*, 1987, 2:68—74.
〔37〕 ________, *ibid.*, 1989, 3:89—101.
〔38〕 王东明，科学通报，33 (1988), 2015—2018。
〔39〕 Wang, D. M. & Gao X. S., *MM-Res. Preprints*, 1987, 2: 75—106.
〔40〕 Wu Wen-tsun, *Some recent advance in mechanical theorem-proving of geometries*, Amer. Math. Soc., 1984, 235—242.
〔41〕 吴文俊，北京 1980 年 *DD* 研讨会论文集，Vol. 2, 1982, 1073—1092。
〔42〕 吴文俊，见上海 *DD*6 研讨会论文集数学部分，1255 (1987), 173—189。
〔43〕 吴文俊，数学学报，3 (1983), 357—360。
〔44〕 Wu Wen-tsun & Wu Tianjio, *MM-Res. Preprints*, 1989, 3: 30—32.

初等几何判定问题与机械化证明*

摘　　要

机械化证明定理，目前值得注意的是 Tarski 关于初等几何与初等代数定理的机械证明法。他以及后来一些研究工作者的方法，大都基于 Sturm 定理的某种推广，这些方法仍极繁复，因之即使使用了计算机，实际上也是难以实现的。本文的目的，在于把定理限制在不牵涉到“之间”关系的情形，应用完全不同的原理给出初等几何定理的机械化证法。这种方法仅用手算即可给出不太简单的定理的证明。

一、问题的提出

Tarski 在 1948 年的一篇经典著作[1]中，解决了实闭域的判定问题，其主要目的之一，是给出初等几何定理的机械化证明。以后 Seidenberg[2]，Robinson[3]与 Cohen[4]等人又给出了 Tarski 定理的不同证法。这些作者还提出了制作某种判定机器以实现这些机械化证明的建议，但是，这种建议离实现之期似颇遥远[5,6]。事实上，只有一些极不足道的初等几何定理，曾在计算机上证明过，例如文献[7,8]。本文的目的是在不考虑

* 本文摘自《中国科学》，1977 年第 6 期。

点的“之间”关系的情况下，给出初等几何判定问题的另一解法。我们的解法，依据了与前面那些作者所使用的完全不同的原理，提供了几何定理的机械化证明，即使颇为困难的定理，仅用纸和笔也往往可实现。依据这种方法编成程序，在计算机上进行计算，也不会发生实质上的困难。

我们考虑的仅限于平面初等几何，虽然所用方法也可应用于其它各种几何。方法的第一步，在于几何问题的代数化。平面中的点，将定义为某一固定域，例如有理数域 R 上的有序数偶。其次是建立一部辞典，把几何关系翻译成为代数形式，不妨把这些看作定义或公理。例如对于若干个点 $A_i=(x_i,y_i)$，不论重合与否，可规定：

A_1A_2 平行于 A_3A_4，

如果 $(x_1-x_2)(y_3-y_4)-(x_3-x_4)(y_1-y_2)=0$，

A_1A_2 垂直于 A_3A_4，

如果 $(x_1-x_2)(x_3-x_4)+(y_1-y_2)(y_3-y_4)=0$，

A_1A_2 的长度平方为 $r^2=(x_1-x_2)^2+(y_1-y_2)^2$，

等等。

如果把基本域 R 易为其它的域，或把点当作其它类型的数组，或改变公理中的代数形式，例如把上面长度平方函数 r^2 改为函数 $r^4=(x_1-x_2)^4+(y_1-y_2)^4$，我们即进入别的几何领域，例如非欧几何，实或复的投影几何，有限几何，等等。但在本文中，将以平面初等几何为限，因为它可以说是具有代表性的。

作为说明，先举一个简单的例子如下。

试考虑下述语句：

(S_n)设 $A_0A_1A_2$ 是一直角三角形，直角在 A_0，记直角边 A_0A_1，A_0A_2 的长为 x_1，x_2，斜边的长为 x_3，则有

$$x_1^n+x_2^n=x_3^n。$$

我们的问题是如何判定语句(S_n)的真伪。我们所用以证明或否定(S_n)的判定算法，应对初等几何中的一切类似的语句(不考虑“之间”的关系)都同样适用。

为解决这一问题，首先注意在语句中出现的那些点，在语句假设所含的条件限制之下，具有普定性质，因此，若将这些点用坐标表示为 $A_0=(x_0, v_0)$, $A_1=(u_1, v_1)$, $A_2=(u_2, v_2)$，诸坐标 v_0, u_1, v_1, u_2, v_2 可视为不定量，而其它坐标与几何度量 x_0, x_1, x_2, x_3 则代数地依赖于这些不定量，依据语句(S_n)中的假设为以下诸代数方程所制约：

$$f_0 \equiv (u_1 - x_0)(u_2 - x_0) + (v_1 - v_0)(v_2 - v_0) = 0,$$

$$f_1 \equiv x_1{}^2 - (u_1 - x_0)^2 - (v_1 - v_0)^2 = 0,$$

$$f_2 \equiv x_2{}^2 - (u_2 - x_0)^2 - (v_2 - v_0)^2 = 0,$$

$$f_3 \equiv x_3{}^2 - (u_1 - u_2)^2 - (v_1 - v_2)^2 = 0。$$

至于语句 S_n 中的结论则相当于等式

$$g_n \equiv x_n{}^3 - x_1{}^n - x_2{}^n = 0。$$

在以后，将以 R 表示有理数域，并取之为基本域。命 A^9 为 R 上的 9 维仿射空间，以($v_0, u_1, v_1, u_2, v_2, x_0, x_1, x_2, x_3$)为坐标，并按上述确定次序排列。于是以上诸方程 $f_i=0$ 在 A^9 中定义了一个 5 维的代数簇，在本例中，在 R 上不可约，并有普点例如($v_0, u_1, v_1, u_2, v_2, \overline{x_0}, \overline{x_1}, \overline{x_2}, \overline{x_3}$)，其中 v_0、u_1, v_1, u_2, v_2 为独立的不定量。于是语句(S_n)的真伪相当于

$$g_n \text{ 在 } V \text{ 上} \equiv 0$$

或

$$g_n \text{ 在 } V \text{ 上} \not\equiv 0。$$

由以上分析可知，一般的判定问题可依下面的形式表达。

问题 在一 R 上以($u_1, \cdots, u_d, x_1, \cdots, x_n$)为坐标的 $n=r$

$+d$ 维仿射空间 A^n 中，设有一代数簇 V，其定义方程是（u_1，…，u_d 是独立的不定量）：

$$\left.\begin{aligned}&f_1(u_1,\cdots,u_d,x_1),=0,\\&f_2(u_1,\cdots,u_d,x_1,x_2)=0,\\&\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\\&f_r=(u_1,\cdots,u_d,x_1,x_2,\cdots,x_r)=0。\end{aligned}\right\}\qquad(\mathrm{I})$$

这一代数簇 V 分裂成若干不可约成份，其维数假定都 $\leqslant d$，其中 d 维成份的总和为 V^*，通常与 V 相合。V^* 的每一成份具有形如 $(u_1,\cdots,u_d,\overline{x}_1,\cdots,\overline{x}_r)$ 的普点，其中 $\overline{x}_i$ 都是域

$$K=R(u_1,\cdots,u_d)$$

上的代数量。另设有一多项式 $g(u_1,\cdots,u_d,x_1,\cdots,x_r)$ 或多项式组 $g_k\in R[u_1,\cdots,u_d,x_1,\cdots,x_r]$。问题是要依据一种算法来判定在 V^* 上是否

$$g\equiv 0(\text{或 } g_k\equiv 0)。$$

在上述问题的提法中，代数簇 V 或 V^* 反映了几何语句中的假设部分。V 或 V^* 将称为该语句的相关簇。若视簇 V^* 为定义于域 $K=R(u_1,\cdots,u_d)$ 上时，其维数为 0。方程组（Ⅰ）的形式说明代数依赖的变量 $x_1,\cdots,x_r$ 应依次地逐一添入域 K。这反映了这样的几何事实：在某些普定的直线或圆上取某些普定的点，又应用作联线，平行线，垂直线或圆，并取直线或圆的交点等等几何手续，依次代数地添入新的点。这些几何作法产生含有 x 与 u 的代数方程，只须用简单的消去法即可化成方程组（Ⅰ）的形式。因为事实上 x 与 u 间的原始方程很少会是超过 2 次的。我们将称变量 u 为参量而 x 为依量。其次注意 V 的诸不可约成份在 R 上的维数都 $\leqslant d$ 这一条件，无非反映了我们所考虑的几何语句都具有确定性质，而把考虑限制于 V^*，则指退化情形将排除在外。这两者事实上都是通常几何定理中隐含，

而不加明述的假设。至于多项式 g 或 g_k 在 V^* 上$\equiv 0$，则是所需证明或否决的语句中结论部分的代数等价式。在以下，我们将称 g 或 g_k 为相应几何语句的判定多项式(组)。

在理论上，Hermann[9]所用的方法，早就足以给出上述判定问题的算法的解答。但是，他的方法如此繁复，不可避免地要引起诸算式的天文数字般的膨胀，以致即使是最简单不过的几何定理，也无法加以证明。相反，我们在下面所用的判定方法，利用了方程组(Ⅰ)的特殊形式，只需用纸笔即可给出很不简单的定理以机械化的证明。

我们的判定方法依据了以下三个定理，证明见第四节。

定理 1 有一算法，据此可将任一确定性几何语句由(Ⅰ)式定义的相关簇 V^*，分裂成 R 上的不可约成份，使每一 d 维成份 V' 视作定义于域 $K=R(u_1,\cdots,u_d)$ 上时，有一形如 $(p_1,\cdots,p_r)$ 的基(即相应素理想的基)，具有以下诸性质：

(1)每一 p_i 是一 $R[u_1,\cdots,u_d,x_1,\cdots,x_i]$ 中的多项式时，对 x_i 的次数 $m_i>0$；

(2)视 p_i 为 x_i 的多项式时，其系数都是 $R[u_1,\cdots,u_d,x_1,\cdots,x_{i-1}]$ 中的多项式，设有公共因子，且对每一 $j=1,\cdots,i-1$，x_j 的次数都小于 m_j；

(3)视 p_i 为 x_i 的多项式时，其主系数是 $R[u_1,\cdots,u_d]$，中的多项式且$\neq 0$；

(4)作为 x_1 的多项式 p_1，在域 $K=R(u_1,\cdots,u_d)$ 中不可约。又对每一 $i>1$，视作 x_i 的多项式 p_i，在依方程 $p_1=0,\cdots,p_{i-1}=0$ 将 $x_1,\cdots,x_{i-1}$ 添入 K 中所得的扩充域上不可约。

显然多项式 $p_1,\cdots,p_r$ 为 V' 所唯一确定到 R 中的因子，我们将称之为构成 V' 的一个优选基(更正确地说是 V' 在 K 上的素理想的一个优选基)。但须注意，诸依量 $x_1,\cdots,x_r$ 依照所给定的次序排列，又注意这一概念事实上为 Gröbner 所引入，被称

为素基，如参阅文献[10]，又可参阅文献[11]中与此密切相关的一个概念：Characteristic sets。

定理 2 设$(p_1, \cdots, p_r)$是相关簇 V 的任一不可约成份 V' 的优选基。有一算法，据此足以对任一 $R[u_1, \cdots, u_d, x_1, \cdots, x_r]$中的多项式 h，确定一等式形如

$$Dh = \sum_{ij=0}^{m_j-1} h_{i_1\cdots i_r} x^{i_1} \cdots x^{i_r} + \sum_{i=1}^{r} A_i p_i$$

使满足以下诸条件：

(1) $D, h_{i_1\cdots i_r}$都是$R[u_1, \cdots, u_d]$中的多项式，且 $D \neq 0$。

(2) A_i 都是$R[u_1, \cdots, u_d, x_1, \cdots, x_r]$中的多项式。

诸多项式 $h_{i_1\cdots i_n}$ 由算法所唯一确定至 R 中的因子，将称为多项式 h 相应于 V'的优选基$(p_1, \cdots, p_r)$的余系式，或简称为相应于 V'的余系式。

定理 3 相关簇是 V 而判定多项式(组)是 g(或 g_k)的一个几何语句，其成立的充要条件是：任意取定依量 $x_1, \cdots, x_r$ 的一个固定次序后，g(或 g_k)对 V^* 的任一不可约成份 V'的一切余系式都恒等于 0。

二、举　　例

在给出这些定理的证明之前(见第四节)，先举数例说明其应用如下。

例 1 对于本文开始所举直角三角形的语句(S_n)，可见相关簇 V 的参量与依量各取为 v_0, u_1, u_2, v_1, v_2 与 x_0, x_1, x_2, x_3 时，在 R 中不可约且有一优选基(p_1, p_2, p_3, p_4)如下：

$$p_1 = x_0^2 - (u_1 + u_2)x_0 + u_1u_2 + (v_1 - v_0)(v_2 - v_0),$$

$$p_2 = x_1^2 + (u_1 - u_2)x_0 - u_1^2 + u_1u_2 - (v_1 - v_0)(v_1 - v_2),$$

$$p_3 = x_2{}^2 - (u_1 - u_2)x_0 - u_2{}^2 + u_1u_2 + (v_2 - v_0)(v_1 - v_2),$$

$$p_4 = x_3{}^2 - (u_1 - u_2)^2 - (v_1 - v_2)^2。$$

由此得

$$g_2 = p_4 - p_2 - p_3。$$

而在 $n > 2$ 时，有

$$g_n \not\equiv 0 \bmod (p_1, p_2, p_3, p_4)。$$

因此语句(S_n)只在 $n = 2$ 时成立，即所谓勾股定理。

例 2 我们的判定方法可用以给出三角恒等式的机械化证明。例如试考虑下述语句：

(S) 若 $A_1 + A_2 + A_3 = 180°$，则

$$\sin 2A_1 + \sin 2A_2 + \sin 2A_3 = 4\sin A_1 \sin A_2 \sin A_3;$$

(C) 若 $A_1 + A_2 + A_3 = 180°$，则

$$\cos 2A_1 + \cos 2A_2 + \cos 2A_3 + 4\cos A_1 C\cos A_2 \cos A_3 = 0。$$

为判定(S)或(C)的真实性，先置

$$\sin A_i = s_i, \cos A_i = c_i (i = 1, 2),$$

$$\sin 2A_i = x_i, \cos 2A_i = y_i (i = 1, 2, 3),$$

$$\sin A_3 = z_1, \cos A_3 = z_2。$$

取 c_1 与 c_2 为参量，而

$$s_1, s_2, z_1, z_2, x_1, x_2, x_3, y_1, y_2, y_3$$

为依量(依上述次序)，则(S)与(C)的相关簇已在 R 中不可约，且有一优选基($p_1, \cdots, p_{10}$)如下：

$$p_1 = s_1{}^2 + c_1{}^2 - 1, \quad p_2 = s_2{}^2 + c_2{}^2 - 1,$$

$$p_3 = z_1 - c_2s_1 - c_1s_2, \quad p_4 = z_2 - s_1s_2 + c_1c_2,$$

$$p_5 = x_1 - 2c_1s_1, \quad p_6 = x_2 - 2c_2s_2,$$

$$p_7 = x_3 - 2z_1z_2, \quad p_8 = y_1 - 2c_1{}^2 + 1,$$

$$p_9 = y_2 - 2c_2{}^2 + 1, \quad p_{10} = y_3 + 4c_1c_2s_1s_2 - 4c_1{}^2c_2{}^2$$

$$+2c_1^2+2c_2^2-1。$$

语句(S)与(C)的判定多项式各为

$$g_s=x_1+x_2+x_3-4s_1s_2z_1,$$
$$g_c=y_1+y_2+y_3+4c_1c_2z_2。$$

验算得

$$g_s=p_7+p_6+p_5+2z_1p_4-2p_3(s_1s_2+c_1c_2)-2c_1s_1p_2 -2c_2s_2p_1$$
$$\equiv 0\operatorname{mod}(p_1,\cdots,p_{10}),$$
$$g_c=-1+p_{10}+p_9+p_8+4c_1c_2p_4$$
$$\not\equiv 0\operatorname{mod}(p_1,\cdots,p_{10})。$$

由此知(S)给出一真恒等式，而(C)则不然。

例 3　试考虑 Simson 线定理，其相应语句为

(S_s)从$\triangle A_1A_2A_3$外接圆上一点A_4，向三角形三边作垂线，其垂足必同在一直线上。

为证此，可简化取外接圆心为$(0,0)$，而半径为r。记$A_i(i=1,2,3,4)$为(x_i,u_i)，而垂足A_j为(x_j,y_j)，$j=5,6,7$。以r，u_1,u_2,u_3,u_4作为参量，

$$x_1,x_2,x_3,x_4,y_5,y_6,y_7,x_5,x_6,x_7$$

为依量，其次序即如上述。

相关 Simson 簇不可约，且有优越基$(p_1,\cdots,p_{10})$如下：

$$p_1=x_1^2+u_1^2-r^2,\qquad p_2=x_2^2+u_2^2-r^2,$$
$$p_3=x_3^2+u_3^2-r^2,\qquad p_4=x_4^2+u_4^2-r^2,$$
$$p_5=2r^2y_5-h_5,\qquad p_6=2r^2y_6-h_6,$$
$$p_7=2r^2y_7-h_7,\qquad p_8=u_{23}x_5-x_{23}y_5+u_3x_2-u_2x_3,$$
$$p_9=u_{31}x_6-x_{31}y_6+u_1x_3-u_3x_1,\qquad p_{10}=u_{12}x_7-x_{12}y_7+u_2x_1-u_1x_2。$$

在上面的方程中，已采用了下述简化符号：

$$\left.\begin{aligned}u_{ji}=u_i-u_j\\x_{jf}=x_i-x_j\end{aligned}\right\}(i,j=1,2,3),$$

$$h_5=u_4x_2x_3-(u_2x_3+u_3x_2)x_4+r^2(u_2+u_3+u_4)$$
$$-u_2u_3u_4。$$

h_6 与 h_7 式与此类似。

判定多项式为

$$g=x_5(y_6-y_7)+x_6(y_7-y_5)+x_7(y_5-y_6)。$$

直截的计算给出

$$g\equiv 0\bmod(p_1,\cdots,p_{10})。$$

这证明了 Simson 语句(S_s)的真确性。

例 4 作为一个比较非不足道的例子，试考虑 Feuerbach 定理，它的相应语句为

(S_F)三角形的九点圆与四个内切傍切圆相切。

试记三角形的三顶点为$(2u_i,2v_i)$，九点圆的中心为(x_1,y_1)，半径为 r_1，而任一内切或傍切圆的中心为(x_2,y_2)，半径为 r_2，又引入变量 z_1,z_2,z_3 各与三角形三边之长相当，于是以 u_1,v_1,u_2,v_2,u_3,v_3 为参量，而 $z_1,z_2,z_3,x_1,y_1,x_2,y_2,r_1,r_2$ 为依量(次序如上)时，相关 Feuerbach 簇 V_F 分裂成四个不可约簇，具有以下形式的优选基$(p_1,\cdots,p_9)$：

$$p_1={z_1}^2-{u_{23}}^2-{v_{23}}^2,\qquad p_2={z_2}^2-{u_{31}}^2-{v_{31}}^2,$$
$$p_3={z_3}^2-{u_{12}}^2-{v_{12}}^2,\qquad p_4=4\Delta x_1-\alpha_1,$$
$$p_5=4\Delta y_1-\beta_1,\qquad p_6=2\Delta x_2-\alpha_2,$$
$$p_7=2\Delta y_2-\beta_2,\qquad p_8=4\Delta r_1-z_1z_2z_3,$$
$$p_9=2\Delta r_2+\gamma。$$

在以上诸公式中，我们采用了下面的简记符号($\varepsilon_i=\pm1$)：

$$u_{ij}=u_i-u_j,\ v_{ij}=v_i-v_j(i,j=1,2,3),$$

$$2\Delta = \begin{vmatrix} u_1 & v_1 & 1 \\ u_2 & v_2 & 1 \\ u_3 & v_3 & 1 \end{vmatrix},$$

$$\alpha_1 = v_{23}v_{31}v_{12} + u_1^2 v_{23} + u_2^2 v_{31} + u_3^2 v_{12} - 2(u_2 u_3 v_{23} + u_3 u_1 v_{31} + u_1 u_2 v_{12}),$$

$$\beta_1 = -u_{23}u_{31}u_{12} - v_1^2 u_{23} - v_2^2 u_{31} - v_3^2 u_{12} + 2(v_2 v_3 u_{23} + v_3 v_1 u_{31} + v_1 v_2 u_{12}),$$

$$\alpha_2 = -[v_{23}v_{31}v_{12} + v_1(u_2^2 - u_3^2) + v_2(u_3^2 - u_1^2) + v_3(u_1^2 - u_2^2) + \varepsilon_2\varepsilon_2 v_{23} z_2 z_3 + \varepsilon_3\varepsilon_1 v_{31} z_3 z_1 + \varepsilon_1\varepsilon_2 v_{12} z_1 z_2],$$

$$\beta_2 = u_{23}u_{31}u_{12} + u_1(v_2^2 - v_3^2) + u_2(v_3^2 - v_1^2) + u_3(v_1^2 - v_2^2) + \varepsilon_2\varepsilon_3 v_{23} z_2 z_3 + \varepsilon_3\varepsilon_1 u_{31} z_3 z_1 + \varepsilon_1\varepsilon_2 u_{12} z_1 z_2,$$

$$\gamma = \varepsilon_1(u_{31}u_{12} + v_{31}v_{12})z_1 + \varepsilon_2(u_{12}u_{23} + v_{12}v_{23})z_2 + \varepsilon_3(u_{23}u_{31} + v_{23}v_{31})z_3 + \varepsilon_1\varepsilon_2\varepsilon_3 z_1 z_2 z_3。$$

注意 $\varepsilon_i = +1$ 或 -1 的选择相应于四个内切或傍切圆,并反映了 Feuerbach 簇的分裂。

语句(S_F)的判定多项式形如

$$g_\eta = (\eta r_1 - r_2)^2 - (x_1 - x_2)^2 - (y_1 - y_2)^2,$$

其中 $\eta = +1$ 或 -1。通过冗长繁复,但直截了当的计算,验得在 $\eta = +\varepsilon_1\varepsilon_2\varepsilon_3$ 时,$g\eta$ 为 $p_1, \cdots, p_9$ 的线性组合,而在 $\eta = -\varepsilon_1\varepsilon_2\varepsilon_3$时则不然。由此知 Feuerbach 语句$(S_F)$为一真实定理,而 $\eta = +\varepsilon_1\varepsilon_2\varepsilon_3$ 则反映了九点圆与相应内切或傍切圆之间的相切方式。注意:在我们问题的提法下"之间"关系并不考虑在内,因而圆的相切方式我们并不重视。因此,我们可以取 $g = g_{+1} \cdot g_{-1}$为判定多项式,同样也可不引入 z_1, z_2, z_3,而在 u_i, v_i, x_j, y_j, r_j 的仿射空间中取定 Feuerbach 簇,这样它在一开始

就将是不可约的。

三、一些引理

为了证明定理作些准备，本节中将考虑一个在基本域 R 上的扩充域 $R(u_1, \cdots, u_d, x_1, \cdots, x_r)$，其中 $u_1, \cdots, u_d$ 在 R 上超越，而 $x, \cdots, x_r$ 在 R 上是代数的，后者的代数扩充则由以下诸方程所确定：

$$\left.\begin{array}{l} p_1(x_1) \equiv p_{10}x_1^{m_1} + p_{11}x_1^{m_1-1} + \cdots + p_{1m} = 0 \\ p_2(x_1, x_2) \equiv p_{20}x_2^{m_2} + p_{21}x_2^{m_2-1} + \cdots + p_{2m_2} = 0 \\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ p_r(x_1, \cdots, x_r) \equiv p_{r0}x_r^{m_r} + p_{r1}x_r^{m_r-1} + \cdots + p_{rm_r} = 0。 \end{array}\right\}(11)$$

我们假定对 $1 \leqslant i \leqslant r$，诸 p_{ij} 都是环

$$p_{i-1} = R[u_1, \cdots, u_d, x_1, \cdots, x_{i-1}]$$

中的多项式，p_{i0} 都是

$$p_0 = R[u_1, \cdots, u_d]$$

中 $\neq 0$ 的多项式，而将 p_i 视为 x_i 的多项式时，p_i 在由

$$p_1(x_1) = 0, \cdots, p_{i-1}(x_1, \cdots, x_{i-1}) = 0$$

等方程所定的域

$$K_{i-1} = R(u_1, \cdots, u_d, x_1, \cdots, x_{i-1})$$

上不可约。这里 K_0 指域 $K_0 = K = R(u_1, \cdots, u_d)$。

以 $\mathscr{P}$ 表示一切指数组 $I = (i_1, \cdots, i_r)$ 的集体，其中 $0 \leqslant i_j \leqslant m_j - 1$。对这些 I 将采用形式记号

$$x^I = x_1^{i_1} \cdots x_r^{ir},$$

任一形如

$$A = \sum a_1 x^I$$

的多项式，其中 a_I 为某一环或域 F 中的元素时，将称为 $F[x_1, \cdots, x_r]$ 中的一个规范多项式。

引理 1 有一算法，足以对任一 p_r 中的多项式 A，唯一地确定一组整数 $s_1, \cdots, s_r \geqslant 0$ 与一组 p_0 中的多项式 A_I，$I \in \mathscr{P}$，使满足以下诸条件：

$(1)_1$ 除去 P_r 上 p_i 间的一个线性组合外，有

$$p_{10}{}^{sI}\cdots p_{ro}{}^{sr}A \equiv \sum A_{Ix}{}^{I};$$

$(2)_2$ A_I 都是 P_0 中的多项式，其系数对 A 作为 P_r 中多项式的系数说来是线性的；

$(3)_3$ $s_1, \cdots, s_r \geqslant 0$ 是使 $(L1)_1$ 与 $(L1)_2$ 成立的最小整数。

证 视 A 与 p_r 为系数在 P_{r-1} 中的 x_r 的多项式，应用除法即可得一最小整数 $s_r \geqslant 0$，使下式成立：

$$p_{r0}{}^{sr}A = Q_r P_r + R_{r-1},$$

其中 Q_r, R_{r-1} 为 p_r 中的多项式，而 R_{r-1} 对 x_r 的次数 $< m_r$。今视 p_r 与 R_{r-1} 为系数在 $R[u_1, \cdots, u_d, x_1, \cdots, x_{r-2}, x_r]$ 中的 x_{r-1} 的多项式，仍用除法可得一最小整数 $s_{r-1} \geqslant 0$，使下式成立：

$$p_{r-1,0}^{s_{r-1}} R_{r-1} = Q_{r-1} p_{r-1} + R_{r-2},$$

其中 Q_{r-1}, R_{r-1} 都是 P_r 中的多项式，而 R_{r-2} 对 x_r 与 x_{r-1} 的次数各 $< m_r$ 与 m_{r-1}。依次进行，可相继得

$$p_{r-2,0}^{s_{r-2}} R_{r-2} = Q_{r-2} p_{r-2} + R_{r-3},$$

$$\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots$$

$$p_{10}{}^{s_1} R_1 = Q_1 p_1 + R_0,$$

其中 R_0 为 $-P_r$ 中的多项式，对 x_i 的次数 $< m_i$，$1 \leqslant i \leqslant r_1$，于是可将 R_0 写成 $\sum A_1 x^I$，而得引理所需符合诸条件的多项式。

引理 2 有一算法足以对任一以 y 为不定量的多项式

$$A = A_0 y^m + A_1 y^{m-1} + \cdots + A_m,$$

其中 $A_i \in P_r$，而 A_0 在 K_r 中$\neq 0$，确定以下两式：

$$HA = B + C_1 p_1 + \cdots + C_r p_r,$$
$$B = B_0 y^m + B_1 y^{m-1} + \cdots + B_m,$$

使满足以下诸条件：

(1) 诸 B_i 都是P_r 中的规范多项式，且 B_0 是 P_0 中$\neq 0$ 的多项式；

(2) C_i 都是$P_r[y]$，而 H 是P_r 中的多项式；

(3) $H \not\equiv 0 \operatorname{mod}(p_1 \cdots p_r)$，

任意满足条件(1)的多项式 B，将称为 p_r 上以 y 为不定量的规范多项式。

证 由引理 1 可得整数 $s_1, \cdots, s_r \geqslant 0$ 与下式

$$p_{10}^{s1} \cdots p_{r0}^{s_r} A$$
$$= A_0' y^m + A_1' y^{m-1} + \cdots + A_m' \operatorname{mod}(p_1, \cdots, p_r),$$

其中 A_i' 都是p_r 中的规范多项式，且 A_0' 在 K_r 中$\neq 0$。设 A_0' 与 $x_{i+1}, \cdots, x_r$ 无关，但确实含有 x_i，因 A_0' 对 x_i 的次数$< m_i$，p_i 对x_i 的次数为 m_i，且在 K_{i-1}中不可约，故由通常的除法程序算法可得 P_i 中的多项式h，k 以及 p_{i-1}中的多项式 A_{00}'，使

$$hA_0' + kp_i = A_{00}',$$

其中 h 在K_i 中$\neq 0$，而 A_{00}'在 K_{i-1}中$\neq 0$。由此得下式

$$hp_{10}^{s_1} \cdots P_{r0}^{sr} A + kp_i y^m = A_{00}' y^m + hA_1' y^{m-1} + \cdots$$
$$+ hA_m' \operatorname{mod}(p_1, \cdots, p_r)。$$

应用引理 1 于 A_i' 与A_{00}'，可得下式

$$h'' A = A_0'' y^m + A_1'' y^{m-1} + \cdots A_m'' \operatorname{mod}(p_1 \cdots, p_r),$$

其中 A_i'' 都是p_r 中的规范多项式，A_0'' 在 K_r 中$\neq 0$，且不含有任一 $x_j, j \geqslant i$。又 h''为P_r 中的多项式，且在 K_r 中$\neq 0$。若 A_0'' 中不再含有任一 $x_1, \cdots, x_r$，即可取 $A_{00}'' y^m + A_1'' y^{m-1} + \cdots + A_m''$ 为引理所需的多项式 B，否则如前进行即得证明。

引理 3 有一算法足以对 $p_r[y]$中的任一多项式

$$A = A_0 y^m + A_1 y^{m-1} + \cdots + A_m,$$

在 K_r 中作因子分解(诸 A_i 都$\in P_r$, 且 A_0 在 K_r 中$\neq 0$, 又 $m \geqslant 2$)。详言之,即可确定形式

$$HA = A' + \sum C_i P_i,$$

$$A' = \bar{A}_1 \cdots\cdots \bar{A}_i。$$

使以下诸条件满足:

(1) 诸 $\bar{A}_j$ 都是 $p_r[y]$中的规范多项式,且在 K_r 中不可约;

(2) H 是 p_r 中的多项式,且在 K_r 中$\neq 0$;

(3) C_i 都是 $p_r[y]$中的多项式。

证 Hermann 在文献[9]中的方法给出了 A 在 K_r 中的因子分解。在通分后即得形式

$$DA = B_1 \cdots B_i + \sum B'_i p_i,$$

其中 B_i, B'_i 都是$P_r[y]$中的多项式, B_i 在 K_r 中不可约, D 是 p_r 中的多项式,而在 K_r 中$\neq 0$。对每一 B_i 应用引理 2 即得所需的形式。又可参阅文献[12, p.130]。

四、定理的证明

至此定理 1—3 可容易地证明如下。

定理 1 的证明.

对于相关簇 V 的定义方程组(I)试视 $f_1(u_1, \cdots, u_d, x_1)$为一 x_1 的多项式,以 $R[u_1, \cdots, u_d]$中的多项式为系数,依据引理 3 的算法,而以 x_1 作为 y, 0 作为 r, 可将 f_1 分解为 K 中的不可约规范因子。任取其一为 $p_1(x_1)$, 命 K_1 为将 x_1 依方程

$$p_1(x_1) = 0$$

添加于域 $K = R(u_1, \cdots, u_d)$所得的扩充域。今视 f_2 为不定量

x_2 的多项式时，必不能在域 K_1 中恒等于 0，否则簇 V 的维数将 $>d$，这与原来几何语句的确定性假设相违。今将 x_2 作为 y，1 作为 r，而应用引理 3 于 f_2，即得一 $R[u_1, \cdots, u_d, x_1, x_2]$ 中的多项式 f'_2 与一式如

$$h_2 f_2 = f'_2 + C_{21} p_1。$$

这里 f'_2 视为不定量 x_2 的多项式时，在 k_1 中 $\neq 0$，且已分解成 K_1 中不可约的规范多项式的乘积。任取一因子为 $p_2(x_1, x_2)$，命 K_2 为将 x_2 依方程

$$p_2(x_1, x_2) = 0,$$

添入 K_1 后所得的扩充域。于是 f_3 为 K_2 上不定量 x_3 的多项式，且不恒等于 0。今再依引理 3 以得一式

$$h_3 f_3 = f'_3 + C_{31} p_1 + C_{32} p_2,$$

其中 x_3 的多项式 f'_3 已分解为规范多项式的乘积，且每一因子在 K_2 中不可约。任取一因子为 $p_3(x_1, x_2, x_3)$，并依方程

$$p_3(x_1, x_2, x_3) = 0,$$

将 x_3 添入 K_2 以得 K_3。如此依法进行，即可得一组子簇 V'，每一 V' 都在 R 上不可约。且在 K 上有一如下类型的定义方程组：

$$\begin{cases} p_1(x_1) = 0, \\ p_2(x_1, x_2) = 0, \\ \qquad \cdots \\ p_r(x_1, \cdots, x_r) = 0, \end{cases} \tag{III}$$

它满足若干明显的条件如第三节所述。

易见，这些子簇构成了整个所给的相关簇 V^*。盖考虑 V^* 在 R 中的任一不可约成份，并设 $(u_1, \cdots, u_d, \bar{x}_1, \cdots, \bar{x}_r)$ 是它的一个普点，其中 $u_1, \cdots, u_d$ 是独立的不定量，而 $(\bar{x}_1, \cdots, \bar{x}_r)$ 代数地依赖于这些 u。因 $(u_1, \cdots, u_d, \bar{x}_1)$ 必须满足方程组

(I)，特别应满足 f_1，故至少需满足 f_1 的一个不可约因子，不妨设即为前面的 $p_1(x_1)$。因$(u_1, \cdots, u_d, \bar{x}_1, \bar{x}_2)$必须满足方程组$(I)$中的其它诸方程以及方程 $p_1(x_1)=0$，故由前面关于 f'_2 的等式知，这一点也将满足方程 $f'_2=0$，因而至少须满足 f'_2 的某一不可约因子，设即为前面的 $p_2(x_2, x_2)$。依次进行可知，$(u_1, \cdots, u_d, \bar{x}_1, \cdots, \bar{x}_r, \cdots)$应满足类型如(III)式的一组方程，为前面所得不可约子簇组中某一个的普点。这就证明了定理 1。

定理 2 与 3 的证明.

定理 2 从引理 1 立即可以得出。定理 3 则从定理 1 与定理 2 得出。

附注.

我们关于初等几何定理机械化证明所用的算法，主要牵涉到一些多项式的运用技术，例如算术运算与简单消元法之类。应该指出，这些都是 12 至 14 世纪宋元时期中国数学家的创造，在那时已有相当高度的发展。详细介绍可参阅钱宝琮的著作[13]。事实上，几何问题的代数化与用代数方法系统求解，乃是当时中国数学家主要成就之一，其时间远在 17 世纪出现解析几何之前。

参 考 文 献

[1] Tarski, A. & Mekinscy, J. C. C, A Pecision Method for Ele mentary Algebra and Geometry, 2nd ed. Berkeley and Los Angeles, 1948—1951 (中译本：初等代数和几何的判定法，科学出版社，1959)。

[2] Seiderberg, A, Annals of Math, 60 (1954). 365—374.

[3] Robinsin, A, Introduction to model Thorg and to the Metamathematics of Algebra, Amsterdem, 1963.

[4] Cohen, P. J,, Comm, Pure Appl. Math, 22 (1969), 131—152.

[5] Robinson, A, Leet Notes in Math, 1975. 498, 14—40.

[6] ——, in Proc. of Tarski Symposium, 1974, 139—152.

[7] Mcharem, J. D. et al. IEE Trans. on Computers, C25 (1976), 773—782.

[8] Keiter, K., IEEE Trans. on Computers, C-25 (1976), 328—334.

[9] Hermann, G., Math. Ann, 95 (1926), 736—788.

[10] Gröbner, W., Moderne Algebraische Geometrie. Wien. 1949.

[11] Ritt, R. F., Differential Algeera, 1950.

[12] Waerden, Van der, Moderne Algebra, Bd. 1, 1930.

[13] 钱宝琮，中国数学史，科学出版社，1964。

几何学机械化方法及其应用*

几何学机械化方法，起源于十二三世纪宋元时期初次出现的几何学代数化，将几何学问题化为多项式问题，以及相伴而生的多项式组消去法。近年来，作者初次利用计算机将这种方法施之于初等几何定理的机器证明与机器发明，稍后又推广之于微分几何，获得了一点成果，介绍如下。

固定一组变量 x_1，…，x_n，以及特征为 0 的域 k，以下所谓多项式，都指在 k〔x_1，…，x_n〕中而言。

多项式组 $TPS=\{f_1,\cdots,f_n\}$ 称为是三角形的，如果可将 x_1，…，x_n 分成两部分，各记为 u_1，…，u_s 与 y_1，…，y_r（$r+s=n$），使诸 f 具有以下形式，并符合诸条件（c_1-c_2）

$$f_1=I_1y_1{}^{m1}+y_1\text{ 的低次项，}$$

$$f_2=I_2y_2{}^{m2}+y_2\text{ 的低次项，}$$

$$\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots$$

$$f_n=I_ny_n{}^{mn}+y_n\text{ 的低次项。}$$

条件是：

（c_1）f_i 中诸 y_i 项的系数都是 k〔u，y_1，…，y_{i-1}〕的多项式，

（c_2）I_i 中 y_i（$j<i$）的次数都$<m_j$，诸 I_i 称为初式。

对于任意多项式 G，可唯一确定一式

---

* 本文摘自《现代数学新进展》（吴文俊主编），安徽科技出版社，1988 年。

$$I_1^{S1}\cdots I_n^{sn}G=\sum Q_i f_j+R,$$

其中 $s_i \geqslant 0$, $Q_i \in k[x_1, \cdots, x_n]$, $R \in k[x_1, \cdots, x_n]=k[u, y_1, \cdots, y_n]$，而 R 中 y_i 的次数都小于 m_i，称为 G 对 TPS 的余式，记作 $R=\mathrm{Remdr}(G/TPS)$。

三角形多项式组 TPS 将称为不可约的，如果诸 f_i 对 y_i 而言在扩域 $k(u, y_1, \cdots, y_{i-1})$ 中都是不可约的，这时 u 的个数 n 称为 TPS 的维数，记作 $\dim TPS$。

对于任一多项式组 PS 以及另一多项式 H，我们将以 Zero(PS/H) 表示，在 k 的任一扩域中使 $PS=0$ 但又不使 $H=0$ 的那些零点的集体。

我们的机械化方法，奠基于下面两个定理，

Ⅰ（Ritt 原理）：

任给一多项式组 PS，可机械地得出一三角形多项式组 TPS（不唯一，称 PS 的特征组），使

(1) $\mathrm{Zero}(TPS/J) \subset \mathrm{Zero}(PS) \subset \mathrm{Zero}(TPS)$；

(2) $\mathrm{Zero}(PS)=\mathrm{Zero}(TPS/J)+\sum \mathrm{Zero}(PSi)$。

其中 J 指 TPS 中诸初式 I_i 的乘积，PS_i 指将 I_i 添入 PS 后的多项式组。

Ⅱ（零点分解定理）：

任给一多项式组 PS 与另一多项式 H，可机械地得出一个分解（不唯一）

$$\mathrm{Zero}(PS/H)=\Sigma \mathrm{Zero}(IRRi/Ri)。$$

其中 $IRRi$ 都是不可约三角形多项式组，而 $\mathrm{Remdr}(Ri/IRRi) \neq 0$。

这两定理都可自然地推广于微分多项式组。

现应用由这两个定理所构成的机械化方法解决下述各问题。

1. 初等几何定理的机器证明与机器发明。

设与所考虑定理假设相当的多项式组为 HYP，求证的结论相当于多项式 $CONC$，依零点分解定理，可机械地求得分解

$$\text{Zero}(HYP)=\sum\text{Zero}(IRRi/R_i)。$$

设 $\dim IRRi=d_i, d=\max d_i$，则定理成立的充要条件是

$\text{Remdr}(\text{CONC}/\text{IRR}i)=0$，$d_i=d$ 时。

$$\sum\text{Zero}(IRRi/Ri),$$

$$d_i<d,$$

提供了使定理可能不成立的退化部分。

由于分解式不甚易求，故在实际应用中使用了 Ritt 原理

$$\text{Zero}(HYP)=\text{Zero}(TPS/J)+\sum\text{Zero}(HYPi)。$$

于是在 $\text{Remdr}(CONC/TPS)=0$ 时，即在 $J\neq0$ 的非退化条件下定理成立。这给出了证明定理的充分条件。在上述条件不满足时，可再进而分解以进行之。

吴文俊、周咸青、王浩集体都曾依据这一方法（主要是充分条件的部分），已经证明并发明了大量定理。对于某一类所谓 Hilbert 型定理，胡森、王东明还编制了特殊的程序，也证明和发明了不少定理。

我们的方法，普遍使用于各种初等几何，包括非欧几何、圆几何等。

2．微分几何定理的机器证明与机器发明。

由于我们的方法依据（Ritt 原理与零点分解原理）可自然地推广到微分多项式，因而定理的机器证明与发明也可自然地推广到微分几何。现在对度量空间与仿射空间中 Bentnard 型曲线偶的定理，已做了较详的分析，其它方面的工作在逐步进行中。

3．未知关系的机械推导。

某些几何量间应有的未知关系，可应用以上方法自然地自

行求出。例如，在一双曲型非欧平面中一个三角形三边之长已知时，我们的机械化方法可自动给出三角形的面积以三边之长表达的形式。双曲型非欧空间中四面体体积的求法，是一个似乎迄今还未获得很好解决的难题，我们的方法提供了一个可以尝试的另一种途径。

类似的问题是大量的，因而我们的方法在应用上有广阔的前景。

4．高次代数方程组求解问题。

对于一个三角形的多项式组，其相应方程组的求解可以归结为单个变量高次方程逐个求解的问题，不论是求复数解、实数解，或整数解，都有众多的现成方法可行。因而零点分解定理提供了一个方法，对于任意一组高次代数方程，可以判断是否有解无解，或有有穷多解和无穷多解，不论事先规定是实解或整数解或否都可。在恰有有限多个解时，可将这些解的全体无增无漏地全部求出。由于整个过程可以形式地进行，直到分成三角形后再进行数值求解，因而误差积累的严重性较小。我们曾以之与国外已有依据多项式处理的两种方法比较，对于他们所提出的例子，我们的方法要优越得多。

由于解高次方程组的问题是普遍出现的，而且实际中广泛使用的超越函数，事实上往往可用多项式来代替，因而我们这一方法的应用前景是很乐观的。

5．因子分解问题。

多变量多项式的因子分解，是计算机科学中一个既重要又极为困难的问题，近年来国外提出了一些较好的方法，其理论依据并不简单。胡森、王东明同志指出，对于整系数多项式在有理数域或任意代数数域中的分解因子，可化为一个求某种特殊形式多项式方程组的整数解问题，由此得到了因子分解的一个新方法。对于某些国外刊物上依据国外的方法在大型计算机

上分解因子的实例，胡森、王东明用他们的方法仅用手算即已很快完成，因而不仅理论简明且方法效率也较国外已知方法要优越。

我国宋元时代所引进的多项式概念与运算以及消元方法，由于其存在的普遍性，在此基础上发展起来的机械化方法，能够获得广泛的应用是自然的。由于表面上看来是超越函数或超越方程定义的概念，有许多可以多项式或微分多项式来取代，因而现代数学中众多领域，都有可能使用这样的方法。

迄今为止，由于设备条件等的限制，我们只使用了微型机或与之相当的小型机，因而进展缓慢，但收获已经不小。今后由于方法与程序的本身改进，特别由于计算机的迅猛发展，内存与计算速度每年成倍增长，可以预见应用计算机来解决所有能机械化的问题，必将收到更大的效果。

计算机的使用不仅对计算机科学，而且对纯粹数学也会提出新的可能从未想到过的课题，计算机的使用与纯粹数学的纯理论研究应该是统一的，而决不是某些人所认为的那样是对立的。

数学的机械化与机械化的数学*

数值计算与定理证明是数学中两项最主要的活动形式。王浩先生[1]在关于机械化数学的论著中,曾列举了这两种活动的若干不同之点。总的说来,计算易而繁,证明妙而难,计算之所以容易,主要是由于计算过程往往已经或易于做到刻板化或机械化。正由于已机械化,才有可能使十七世纪时巴斯喀用齿轮传动造出了加法机,并由莱布尼茨改进成乘法机,到了现代更使已经机械化了的多种计算过程通过电子计算机实现了自动化。与之相反,即使叙述颇为简单的初中几何证明题,也往往使许多大几何学家棘手。如何使定理证明化难为易,甚至不惜弃简就繁,乃是机械化证明的目的所在。为此提出下面的问题:定理证明是否也可像数值计算那样机械化,进而通过电子计算机以实现定理证明的自动化。

这一问题并非完全是现代的产物。事实上,早在17世纪莱布尼茨时就有机械化证明的设想。只是直到19世纪末及以后由于希尔伯特及其追随者们建立并发展了数理逻辑,这一问题才具有明确的数学形式。又由于40年代电子计算机的出现,才使这一设想有了现实的意义。几十年来,数理逻辑学家为这一问题付出了巨大劳动,但是,他们所得到的成果基本上是反面的,即所谓不可判定性,例如有:

哥德尔定理 初等整数论的定理证明不可能机械化。

---

* 本文摘自《系统工程论文集》,1981年。

值得称道的正面的结果则有

塔斯基定理 T[2]　初等几何(以及初等代数)的定理证明可以机械化。

塔斯基等还因之提出制造所谓判定机器(也就是证明机)的设想。然而他们的方法与设想都不是切实可行的。1976 年 IEEE 刊物上有一篇关于美国机器证明试验的报道[3、4],用塔斯基方法于电子计算机只能证明一些近于同义反复的儿戏式"定理",例如,"若 $P_1, \cdots, P_5$ 五点中 P_1, P_2, P_3 又 P_1, P_2, P_4 又 P_1, P_2, P_5 各在一条直线上,则 P_3, P_4, P_5 也在一条直线上。"总之,直到目前为止,国外对于机械化证明的尝试处于胶着状态,劳而少功,实质上是失败的。

但是,甚至在 1899 年出版希尔伯特的经典著作《几何基础》[5]中,就有着耐人寻味的正面的机械化结果,只是从来没有人注意过,也许包括希尔伯特自己在内。该书向来被认为是公理化的典范,但实质上希尔伯特却指出了(也许是不自觉的)如何从公理化通过代数化到达机械化的道路。作为该书主体部分前六章的最后一个定理(第八版定理 62),即可解释成为下面的:

希尔伯特机械化定理 H　初等几何中只涉及从属与平行关系的定理证明可以机械化。

这一定理自然是塔斯基定理 T 的一个特例。但与塔斯基定理不同,希尔伯特的机械化证明方法是切实可行的。甚至只用手算就可证明颇为复杂非不足道的(只涉及从属与平行关系的)定理。依据它也不难编成程序在计算机上实现。

现将希尔伯特的机械化方法大致说明如下。

希尔伯特把初等几何的公理分成五类:(Ⅰ)从属公理,(Ⅱ)次序公理,(Ⅲ)全合公理,(Ⅳ)平行公理,(Ⅴ)连续公理。现假定空间的第Ⅰ,Ⅱ类公理与第Ⅴ类公理成立,则在直线上任意取定两点 0 与 Ⅰ,直线上的点之间即可引进加法与乘法以得一满

足通常的结合、交换与分配等定律的数系统,其中 0 与 I 各与 0 与 1 相当,但乘法交换律除外。如果我们的几何是所谓巴斯喀几何,即巴斯喀定理成立,则这一数系统也将满足乘法交换律,依据希尔伯特的证明,如此所得的数系统与直线以及上面 0 与 I 的选择无关,即唯一确定至一同构。在空间中任取三个不同平面的共点直线以公共点作为这些直线上建立数系统时的 0 点,则依通常方法即可建立一坐标系统,点、直线、平面等几何元素可用数系统中的数或数之比来表示。把它们统称为几何数标,则一切几何元素间的从属或平行关系都可通过它们的几何数标的多项式等式关系来表达,而多项式中每一数标的次数都是一次。这就是希尔伯特从公理化出发以到达代数化的具体步骤。

一个通常的几何定理,无非是说如果某些几何元素满足若干几何关系(假设),则它们必将满足某一几何关系(终结)。任取一坐标系并把它们都用数标表示,则其中某些数标可以是任意的,称之为参数,另一些则须受假设中几何关系的约束而不再任意,称之为束数。又由于定理中的几何元素往往是通过联线,作平行线,相交等手续逐步作出,因而可将这些数标排成某一次序,并将参数与束数各依次记为 $u_1, \cdots, u_d$ 以及 $x_1, \cdots, x_r$。经过简单化约即可将假设表述为下面的方程组

$$A_i x_i + B_i = 0, \ i = 1, \cdots, r, \tag{1}$$

其中,A_i, B_i 都是 $u_1, \cdots, u_d$ 与 $x_1, \cdots, x_{i-1}$ 的多项式。同样终结可表达为

$$g(u_1, \cdots, u_d, x_1, \cdots, x_r) = 0, \tag{2}$$

其中,g 也是多项式。于是原来的几何定理无非是说:

方程(2)是方程组(1)的自然结果。

这一断言是否成立是容易验证的。事实上,从式(1)经过逐步约化可依次得

$$x_i = P_i / Q_i, i = 1, \cdots, r, \tag{3}$$

其中,P_i, Q_i 都是 $u_1, \cdots, u_d$ 的多项式。代入 g 中,得

$$g = g\left(u_1, \cdots, u_d, \frac{P_1}{Q_1}, \cdots, \frac{P_r}{Q_r}\right)。$$

由于 u 都是参数,g 之是否为 0 即可直接从计算得出。

显然所有的过程都可机械地进行,这就是希尔伯特的机械化证明方法。这一结果也可概括为

希尔伯特机械化定理 P 巴斯喀几何的定理证明可以机械化。

由于巴斯喀定理可利用全合公理证明,因而定理 P 自然包括了希尔伯特机械化定理 H。

在 1976 与 1977 年之交,作者开始从事定理证明的机械化问题,获得了下面的结论[6]。

定理 1 初等几何中只牵涉到从属、平行与全合关系的定理证明可以机械化。

我们机械化证明方法实质上与希尔伯特的方法不约而同。例如与前同样选择坐标系,将诸几何元素的数标分成参数与束数两类并排成次序为 $u_1, \cdots, u_d$ 与 $x_1, \cdots, x_r$。所不同者是与假设部分相当的方程不再对每一 x_i 是线性的,而有较一般的形式:

$$P_{i0}x_i^{mi} + P_{i1}x_i^{mi-1} + \cdots + P_i^{mi} = 0, i = 1, \cdots, r, \tag{4}$$

其中,诸 P_{ij} 都是 $u_1, \cdots, u_d$ 与 $x_1, \cdots, x_{i-1}$ 的多项式。记基域添加 $u_1, \cdots, u_d$ 后的域为 K_0,则在通常可称为"不可约"的情形,在 K_0 中 $P_{10} \neq 0$,且$(4)_1$ 不可约。记 K_0 添加满足$(4)_1$ 的 x_1 后的域为 K_1,则在 K_1 中 $P_{20} \neq 0$,且$(4)_2$ 不可约,依次类推。记$(4)i$ 的左边为 f_i,则任一多项式 $g(u_1, \cdots, u_d, x_1, \cdots, x_r)$用除法可得

$$P_{10}^{s_1} \cdots P_{r0}^{s_n} g = g^0 + \sum_{i=1} h_i f_i,$$

其中，g^0 对每一 x_i 的次数都 $< m_i$，称为 g 对 f_i 的余式。于是定理结论式(2)成立的充要条件是余式 $g^0=0$。在一般情形，即 f_i 在相应域中非不可约时，则可先将 f_i 分解因子后进行，而分解因子的手续也是可以机械化的。

显然定理1包括了希尔伯特的定理H，但仍包括于塔斯基的定理T之中。然而，与希尔伯特定理相似而与塔斯基定理不同，我们的机械化方法是切实可行的，即使用手算也可证明例如富尔巴赫那样艰深的定理。据此编成程序以在计算机上实现，并无实质上的困难。数学所陈东岳等同志以及美国王浩先生都曾在电子计算机上实施过。作者本人也曾在国产长城203台式机上实施过。三者都曾证明西姆森线那样非不足道的定理。我们的经验是：即使是一台小型台式计算机，只须容量足够大（现在 *HP*9835 或 9845 型台式机的内存已可达到 256*K* 以至 400 余 *K* 字节，这里每字节合 8 字位），应可证明极为繁复艰深的定理，而计算时间完全可以接受（例如几个或几十个小时）。

我们还可对所需计算量作出理论的估计，如下面的

定理 2　初等几何中只牵涉从属、平行与全同关系且“不可约”的定理证明，其计算量满足下面的关系：

计算复杂度≤定理复杂度·几何复杂度。

上述定理中“不可约”即前面提到通常的情形，往往从定理叙述的几何内容即可一望而知。计算复杂度依斯特拉逊定义[7]，定理复杂度与几何复杂度可用具体的公式表达，也可从“图形”直接读出。

1977与1978年之交，作者又将初等几何的结果推广至初等微分几何，即[8]。

定理 3　初等微分几何中凡可用微分多项式等式关系来表达的定理的证明可以机械化。

这里定理证明的机械化方法依赖于里奇埃与里脱关于微分

方程与微分代数的理论与方法[9]，比初等几何的情形要复杂得多。但据之以编成程序以在计算机上实现仍非难事，而且粗略估计一具 *HP* 型的台式计算机应该已足够了。

上面一些结果的相互关系可用图形示意表达。

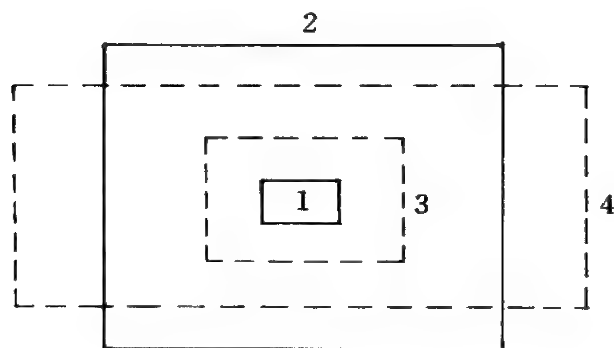


图 1

直到现在，我们还没有找到一种切实可行的方法来实现整个初等几何定理证明的机械化。这牵涉到希尔伯特著名的第17问题。这一问题虽已在1926年为阿丁所解决[10]，但他的解决方法只是存在性的。阿丁曾提出寻找构造性解决方法的问题，经过四十多年到近年才出现这样的方法[11]，但仍不是切实可行的。而且希尔伯特第17问题的提法与我们的要求还有距离。总之，这是一个理论的探索性问题，解决似非易事。

另一方面，我们的结果适用于各种克莱因型的几何，而不只是初等几何而已。

我们的结果还表明，尽管某一数学领域整个说来是不可能机械化的，但并不排除其中一部分可以机械化。如何发现这样一些可以机械化的部分领域，提出切实可行的机械化方法，又是一项高度理论的探索性问题，只有对该领域有深邃认识才有解决的希望。

最后应该指出：我们从事机械化定理证明工作获得成果之前，并不了解塔斯基的已有工作，更没有想到与希尔伯特《几何

基础》会有任何关系。事实上，我们是在中国古代数学的学习与启发之下提出问题并想出解决办法来的。其根本原因是：中国的古代数学基本上是一种机械化的数学，而几何的代数化又是宋元时期我国数学的主要成就。我们的研究工作还只是一个开端。如何继续发扬中国古代传统数学的机械化特色，对数学各个不同领域探索实现机械化的途径，建立机械化的数学，则是本世纪以至可能绵亘整个21世纪才能大体趋于完善的事。

参 考 文 献

[1] Wang Hao, IBM J. Res. Dev. 4(1960), 2—22.

[2] Tarski, A., A decision method for elementary algebra and geometry, 1951 (陆钟万译，1959)。

[3] Kister, K., IEEE Trans on Computers C-25(1976), 328—334.

[4] Mcharem, J. D., et al., ibid, 773—782.

[5] Hilbert, D., Grundlagen der Geometrie, 8ed., 1956(江泽涵等译，1958)。

[6] 吴文俊，《中国科学》(1977), 507—516。

[7] Von V. Strassen, J. & M. V. 78(1976), 1—8.

[8] 吴文俊，《科学通报》(1978), 523—524。

[9] Ritt, R. F., Differential algebra, 1950.

[10] Artin, E., Hamb, Abh., 5(1927), 100—115.

[11] Pfister, in Mathematical developments arising from Hilbert problems, vol. 2(1976). 483—490.

《几何定理机器证明的基本原理（初等几何部分）》的导言*

几何学起源于人们观天测地这一类实践活动，这一点似已成为定论，无可置疑。但不同地区的人们在总结各种经验，综合上升为理论，最后发展成为一门科学的过程中，却有着不同的方式、方法与途径。恩格斯说过，"纯数学的对象是现实世界的空间形式和数量关系"。古希腊时代，对待几何学就有两种不同的方法：一种可以欧几里得的《几何原本》为代表，把数量关系完全排除在外，而单纯追求各种几何事实间的逻辑关系，以此建立几何公理体系，成为数学中演绎推理方法的典范；另一种可以阿基米德的有关著作为代表，着重研究几何图形的数量特征或其量度，诸如圆周率、球面面积以及抛物线弓形面积的计算等等。尽管这二者各具特色，风格殊异，体现了几何发展中的两种不同体系，但都为数学发展作出了巨大贡献。

我国几何学的发展则是始终与数量关系形影不离的。诸如面积、体积的量度以及关于勾股定理的应用，在我国古代几何中一直占据着中心位置。那时我国没有陷入烦琐的成堆公理之中，而是提炼出几条一般原理，并以此作为广泛应用与推理论证的依据。例如，出入相补原理、体积理论中的刘徽原理以及

---

* 本文摘自《几何定理机器证明的基本原理（初等几何部份）》，（吴文俊著）科学出版社，1984 年。

与 16 世纪欧洲出现的 Cavalieri 原理相当的刘祖原理等。所谓阿基米德与欧几里得的两种不同体系，在我国古代几何中兼而有之。勾股定理自然地引起平方根的计算问题，而我国古代求平方根和立方根的方法，其步骤就是以出入相补原理为几何背景逐步索骥而得。由平方根和立方根的计算逐步发展为二次以至高次方程的解法及其理论，最后导致天元术的发现以及几何的代数化、多项式的运算与消去法等。本书所阐述的几何定理证明的机械化问题，从思维到方法，至少在宋元时代就有蛛丝马迹可寻。虽然这是极其原始的，但是，仅就著者本人而言，主要是受中国古代数学的启发。

在古代几何的发展中，由于对数量关系的不同认识而引起了各种倾向，直到 17 世纪解析几何的出现才得到了统一。这使得几何的问题可以用代数的方法来处理，以致可以避免欧几里得那种几何综合证法所需要的高度技巧。虽然几何综合方法以其直观引人入胜，且曾在 19 世纪有过一段复兴的时期，但终究难与解析几何的代数方法相匹敌。现代几何学的许多最重要的领域，例如微分几何与代数几何，往往一开始就假定了一个数系统（一般是一个数域，甚至是特殊的实数域与复数域），由此构成仿射空间或投影空间，然后用坐标或以函数与导数间的代数关系式来引进各种几何图形，如曲线、曲面及其几何关系等等。纯正的从公理出发的综合方法只占据着一个较小的、主要是初等的角落，且应用这种方法往往是出于对几何学基础上的考虑才对公理间的逻辑关系进行细致的分析，并不是为了丰富几何学的具体内容而叠瓦添砖。在数的基础上建立起整个几何学的大厦，看来已是大势所趋。

然而，尽管几何学可以从纯粹数量关系的形式上建立起来，但总不能不考虑几何的直观背景与来源，也不能不考虑它的基础问题。如何从原始的现实形象提炼出一套几何学的公理

系统（或若干原理），又如何从公理系统发展到坐标系统，使代数方法得以在几何中发挥作用，这些都是应当说明的问题。对于称为欧几里得的那种常见常用的几何学，在 Hilbert 的《几何基础》（见 Hilbert〔2〕）这一经典著作中已经作了详尽的论述。我们容易认为，从几何公理引入坐标建立解析几何是轻而易举的事情。但这是由于我们运用了关于实数系统与初等几何的全部知识所产生的错觉。事实上，这一过程是颇为曲折与艰难的，我们仅从 Hilbert 的书如何依据公理系统引入数系统再引入坐标的过程就可以看清楚。在此过程中，Hilbert 所称的两条 Desargues 公理以及一条 Pascal 公理扮演了极其重要的角色，而且就象本书第二章和第六章所指出的那样，投影几何与仿射几何数系统的引入有很大的区别。虽然这两者都只依赖 Desargues 公理（以及一些最简单的关联公理等）就可在各直线上引入彼此同构的数系统，但对于仿射几何，这些数系统之间有唯一确定（Canonical）的同构关系，而投影几何则只能在所谓 Pappus 公理成立的假定下才能唯一确定这些同构。一般的几何学著作，似乎都忽略了这一点，其原因之一可能是偏重于欧几里得传统，着眼于公理间的逻辑关系，而置数系统与坐标的引入于次要的地位，或直接从数系统出发建立几何学，而不再考虑几何对象的现实来源与公理基础。把公理系统与数量关系之间的沟通作为探讨的主题，迄今为止，Hilbert 的书仍是最具代表性的著作。由此还可看出，即使是人所熟知的那种投影几何与仿射几何，其坐标的引入也决非如通常想象的那样简单。至于奠基于种种不同的公理系统之上的各种不同几何，其数系统与坐标系统引入的困难程度，就更不难想象了。

不仅如此，即便从公理系统出发，到达了坐标系统并建立了相应的解析几何，使几何定理的证明完全化成纯代数问题，也并不见得就一定容易。首先，由于解决这些代数问题的计算

量往往过大，使人望而却步；其次因为代表几何关系而出现的那些代数关系式往往杂乱无章，使人无所措手。从这些杂乱无章的代数关系式中要找出一条途径，以到达要证明的那些关系式往往需要高度的技巧。这只要翻阅一下过去的大量著作，例如 Salmon〔1，2，3〕等书就清楚了。现在，由于计算机的出现，对繁杂的计算已经有了有效的处理办法。因此，如何把杂乱无章的代数关系式整理得井然有序，使计算机得以发挥其威力，便成为整个问题的关键所在。

至此，几何定理的证明问题可以分成下面三个主要步骤：

第一步，从几何的公理系统出发，引进数系统与坐标系统，使任意几何定理的证明问题成为纯代数问题；

第二步，将几何定理假设部分的代数关系式进行整理，然后依确定步骤验证定理终结部分的代数关系式是否可以从假设部分已整理成序的代数关系式中推出；

第三步，依据第二步中的确定步骤编成程序，并在计算机上实施，以得出定理是否成立的最后结论。

我们称第一步为几何的**代数化**与**坐标化**，第二步为几何的**机械化**，至于第三步能否使用计算机作最后验证，完全依赖于第二步机械化之是否可能。由于计算机只能识别有限的事物，因此一个先决条件乃是第二步中的代数关系式必须都以有限的形式出现。所以，如果第二步中的代数关系式牵涉到连续、极限一类概念，甚至以超越函数的形式出现，即排除了使用计算机的可能性。相反，如果这些代数关系式都是以多项式的形式出现，且其系数都是整数，则对于计算机的使用只是依据第二步中所确定的步骤编制程序的问题，不会有任何实质的困难。如果一门几何可以找到这样三步（事实上只要前面两步即可）来完成定理的证明，我们就说这门几何**可以机械化**，并把可以机械化的结论称为**机械化定理**。机械化定理是否成立，依

赖于第一、二两步能否实现，而这两者完全是纯理论性的问题。

一门几何能否机械化，并不显然。相反，由于通常的各种几何广泛使用了连续、极限一类概念，尤其是微分几何离不开函数与微分，因而从表面上看似乎不能使用计算机进行证明。但事实上并非如此，其道理颇为深奥。按 Hilbert 的《几何基础》其重要的一点是说，通常几何的基础完全可以排除连续这样一类公理。在该书俄译本 Rashevsky 所作的序言中，就曾明确指出了其中心思想，实质上就是几何学的发展可与连续公理无关。在 Hilbert 的另一篇关于 Bolyai-Lobachevsky 非欧几何的文章（Hilbert〔2〕）之末，也明确指出了这一点。正因为如此，几何的公理、定理与证明，实质上都可以用有限次的构造步骤来叙述以至完成。这也就提供了几何定理证明有可能机械化的依据。就著者所知，最早的一条真正的机械化定理即见于 Hilbert 的《几何基础》一书，本书第三章即致力于这一机械化定理的证明，并称之为 Hilbert 机械化定理，参见 Wu Wen-jun〔4〕。

同样的论点也适用于微分几何。如果抛开几何的实质，只从形式逻辑关系来考虑（定理的证明要求这一点已经足够），在微分几何的各种概念与定理的叙述中出现的那种函数与导数完全可以形式地来对待，而不必考虑其是否与真正的连续性或极限过程有关。因此，即使是微分几何的定理证明，也存在着通过有限次形式上的构造步骤借助于计算机来进行定理证明的可能性。事实上，情况也确实如此。我们已经通过计算机证明了微分几何的某些定理，而且在微型计算机上也做过一些试验，参阅 Wu Wen-jun〔2，3〕两文。

我们把几何按其是否涉及微分而粗略地分成两大类，一类是微分几何，另一类则是所谓的**初等几何**。为方便，现改称通

常称为欧氏几何的那种初等几何为**常用几何**。著者将陆续阐明各种几何机械化的理论与方法。本书只论述初等几何机械化的基本原理，在另一本与此并行的姐妹篇《几何定理机器证明的理论、方法与实践（初等几何部分）》中，将阐明这一机械化理论与方法在计算机上的具体实施，其中包括程序的编制，计算量的估计，具体定理的证明，新定理的发明以及几何的理论和方法对计算机使用效率的改进与各种应用等等。在本书以后的有关书籍中则将致力于阐述微分几何的机械化问题以及各种有关的理论问题。

本书共分六章。前两章是关于几何机械化的预备知识。作为实例，我们集中讨论常用几何，即通常所称的欧几里得几何，并详细说明从公理出发建立坐标系统的过程。为简便起见，只讨论平面情形；又为了讨论有所遵循，一切都围绕 Hilbert 所提出的五类公理进行（这五类公理是关联公理、次序公理、全合公理、平行公理与连续公理）。Hilbert 在建立常用几何时摈弃了连续公理，但次序公理和关联公理则贯穿在他的整个体系中，甚至全合概念与全合公理本身都不能独立于次序概念之外来叙述。如何不依赖于次序概念与次序公理而较早地引入垂直、全合以及其它度量概念，似乎为 Hjelmslev 首先提出，并一直延续到本世纪五六十年代，参阅 Bachmann〔1〕及 Klingenberg，Lenz，Reidemeister，Schülte，Sperner，Winternitz 等的著作。他们的着眼点似在研究反应一些几何主要概念的不同公理组之间的逻辑依存关系。与此相反，本书的目的在于几何的机械化，而不在于几何的公理化。对于公理的独立性与相互依存问题不是我们的旨趣所在。但是就机械化这一主题而论，在机械化的理论与方法上有无次序公理却大不相同。此外，在一些现代几何，例如代数几何中，主要考虑的是复数域或特征为 0 的任意数域，根本不存在任何次序关系。即使在

有次序关系的几何中，真正涉及到次序关系的定理也并不占主要地位。凡此种种，说明在从公理出发建立几何时，次序关系有必要退居较次要的地位。为此，在前两章建立常用几何的过程中，次序关系尽量移后，而逐步引入各种无序几何与有序几何。在引入几何的数系统以完成前面第一步所说的代数化与坐标化时，虽然仍象 Hilbert 一书那样，主要依据以 Desargues 与 Pascal 命名的两条公理，但已不再象 Hilbert 那样把次序公理掺杂其间。此外，我们又添加了一条无限公理，借以排除一切有限几何。因为从机械化的角度来看，有限几何可以机械化是显然的。

本书的后四章致力于几何的机械化问题。传统的欧几里得证明定理的方法要求对个别的定理寻求个别的证法，而且每一证明总是要求某种新的、往往是奇巧的想法（参阅 Kline〔1〕的 307—308 页，中译本的Ⅱ. 8）。本书关于几何定理证明的机械化寻求的是一般方法，它不仅适用于个别的定理，而且适用于整个某一类型的定理，甚至可以说是某一种几何的所有定理。只要依照书中所述的方法机械地进行，在有限步之后，就可对整个一类定理得到统一的证真或证伪，而无分难易，要做到这一点，必须通过以数量关系为主的代数方法来实现，而几何的代数化乃是关键性的一步。这正好符合 Descartes 采用代数方法把推理程序机械化以致减小解题工作量的想法（见 Kline〔1〕），也正象早期的几何代数化天元术的创立者之一，我国元代朱世杰（13 世纪）在《算学启蒙》（1299）中说“以天元演之，明源活法，省功数倍”。总之，如果真能做到有效的机械化，则不妨借用 Chasles［1］269 页上的一句话：“为几何巨厦添砖加瓦，从此就用不着天才那样的人物了”。

并不是每一种几何或任意一类定理都能找到机械化的证法。对于确实可以进行机械化证明的定理，我们提出三种不同

的类型，它们可以用三种机械化方法来证明，并将分别在三、四、五章中讨论。每一种类型都须假定第一步的代数化与坐标化已经完成，而且可把几何定理的证明问题化为一些代数关系式的处理问题。第一种类型定理的特征是假设部分的代数关系式对于某些特定变量都必须是线性的。这类定理包括所谓一类构造型的纯交点定理，其证明的机械化方法也见 Hilbert [1]。为此我们称这一结果为 **Hilbert 机械化定理**。第二种类型定理的特征是假设与终结部分的代数关系式都可以多项式的方程来表示。这种类型定理的机械化证明方法在 Wu Wen-jun [2] 一文中首先提出，并在多种计算机上实施过。第三种类型定理的特征是假设与终结部分可以是任意的多项式等式或不等式，但其系数必须在一实闭域中，因而原来的几何必须有次序关系。这种类型定理的机械化证法为 Tarski 在 1950 年 [1] 一文中给出，因而我们称相应的结果为 Tarski 机械化定理。上述三种机械化定理各有其适用的范围，特别是第二与第三两种互不包含。我们的机械化定理不适用于终结含有次序关系的那种定理的证明，而 Tarski 的机械化定理则只适用于几何数域为实闭域而不能适用于几何数域为复数域或其它一般数域的情形。从适用的角度讲，三者（各以 Ⅰ，Ⅱ，Ⅲ 表示）的相互关系如图 1 所示；就方法而论，则三者各异。因而在所属范围互相重叠的那些部分，其定理可用几种不同的机械化方法来证明。就证明的效率来讲，Ⅰ 最高，即使用手算也可以证明颇不简单的定理。Ⅱ 次之，用手算虽难以见功，但若使用计算机，即使是一台微型的计算机也可以证明很不简单的定理。至于 Ⅲ，则效率颇

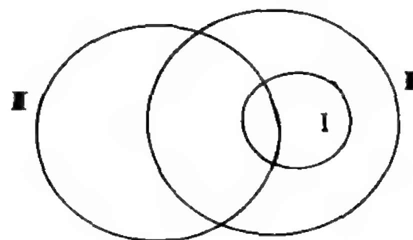


图 1

低，迄今为止，还没有听说过据之以证出什么稍有意义的定理来，即使用大型快速的计算机也是如此。

由于 Hilbert 的机械化方法使用范围过狭，而 Tarski 的机械化方法效率又过低，故在本书的第三、五两章对此分别作了简短的介绍。重点是第四章。这一章的机械化方法在理论上需要代数几何的帮助。虽然现代的代数几何是当前数学中最活跃的分支，但是却不合我们的需要。原因是现代的代数几何几乎纯粹是存在性的，而对于机械化的要求来说，所提供的手段必须是构造性的，只有这样才有可能将相应的步骤在计算机上逐一实现。所幸的是，Ritt [1，2] 两书中早已发展了这种代数几何的构造性理论，恰好符合我们的需要。只是 Ritt 的论证使用了分析的方法，而连续、极限等概念与机械化证法不相容，所以必须进行适当的改造。第四章中的部分内容即致力于此。其中如多项式组的约化整序等概念与方法都出自 Ritt，只是在叙述上有所不同。Ritt 方法还用于微分几何定理的机械化证明，其简略的叙述见 Wu Wen-jun [2，3] 诸文。对此，著者在有关的书籍中将有更详尽的论证。

应用这几章的方法来考察几何，则可证明在第一、二章中通向常用几何的过程中所出现的各种无序的与有序的几何，从所谓无序 Pascal 几何起，其定理的证明都可机械化。在第六章中，我们同样考察了与常用几何有别的一些几何，如投影几何、双曲型与椭圆型非欧几何、Möbius 与 Laguerre 的两种圆几何学等，证明了类似的机械化定理。因为这些几何从公理化到坐标化的过程颇为冗长，故不能象第一、二章那样再作详细论证。

虽然本书展示了不少可以机械化的几何，但决不是任意一种几何都可以机械化，关键在于 Pappus 或 Pascal 公理是否成立，或几何数系统中的乘法交换律是否成立。具体说来，我们

有下面的推测：

推测 Desargues 几何不能机械化。

但我们未能证明这一推测，数理逻辑的专家们可能会作出正确的答案。

自著者在 1976 年和 1977 年之交发现本书第五章所示几何定理的机械化证明方法以来，已经荏苒五年有余。这期间，著者受到了许多鼓励，有时颇出本人意料之外的热心帮助。由于人数太多，无法一一列举，现只略举三人。

在理论上著者深切感谢胡世华同志和王浩先生的鼓励与支持。著者关于定理证明机械化的思想和方法与长期形成的传统思想方法相抵触，如果当时不是胡世华同志从数理逻辑的角度伊始就予以首肯，中途夭折也不是不可能的。至于王浩先生，他在机器证明上突破性的成就早已脍炙人口，而其有关机器证明的一些精辟论点，更是发人深省。诸如以量的复杂取代质的困难，以及基础机证与特例机证应有所区别等论点，都使著者在研究过程中深受启发。

孟繁书同志曾设法为著者在长城 203 台式计算机上进行实践，提供了许多方便。由于在那台计算机上进行了几个月的实践，才使著者有决心设法购取一台较现代化的台式计算机进行多种试验。这些试验说明我们的方法可以付诸实践，而不致停留在纯理论上，成为纸上谈兵。因此，特向孟繁书同志致以最深切的谢意。

复兴构造性的数学*

非构造性观点在现代数学研究中普遍流行。这种观点往往主要考虑对象的一些性质，如存在性、可能性等问题，不大关心如何求出解答、或将能行的方法予以有效的实现。应用上对构造性数学要求非常迫切。一个工程师对于方程解的存在唯一性不会有太多的兴趣，而更关心一些典型的特解，或利用微扰方法找出近似解。机器定理证明向数学提出许多构造性的问题，例如将代数簇如何分成不可约分支，把一正定多元多项式如何表示成为有理函数的平方和等。这些问题在非构造性观点下被搁置多年，目前尚无有效的处理方法。历史上，中国古代数学基本上是构造性的。在西方，非构造性观点从上世纪末才逐渐盛行。实际研究中有许多问题，一时难以给出构造性的处理，因而首先研究存在性、可能性等有关问题，但最终应是构造性的。值得注意的是，近来由于各种原因的促进，构造性观点的抬头有了一些明显趋势。

举一些例子。

例 1 线性联立方程组。按非构造性的观点，仅考虑有没有解，解空间的性质等，构造性观点还要求给出求解方法。50年代计算机的发展促使人们认真考虑这一问题，Cramer 办法形式漂亮，实际上并不可行，Gauss 消去法才是行之有效的。在中国，很早就出现了完整的消去法，于 50～100B. C 成书

---

* 本文摘自《数学进展》，1985 年第 14 期。

的《九章算术》卷八方程术中就详细记述了这种方法。其中列举了若干例题，第一问完整地给出了方法的步骤，其它各问则指出了不同的要点，且已出现正负数的概念。近来，我们用机器证明中的整序原理，对高次联立方程组提出类似的三角算法。

例 2 方程论。其中代数基本定理非常重要，但它是非构造性的，进一步地给出求解方法按照 Galois 理论就必须有限制，只能求近似解，局限在实数范围内西方在 19 世纪有 Horner 方法。中国发展到宋代，秦九韶(1247 年)已提出完整的增乘开方法。近几年 Smale、Kuhn 等人分别根据前人的工作创造了一些方法，进一步提出概率可行性、同伦算法等概念，对算法的效率提出较满意的分析。

上面提到的 Galois 理论也是非构造性的。一般的高次($n \geqslant 5$)方程若仅用加减乘除、开根号等运算是不可解的，因为对应的 Galois 群不可解。正因为没有有效的方法求 Galois 群，所以给了一个高次方程，我们还是无从知道它是否可解。从另一角度看，如果扩大求解的工具(例如包含椭圆函数等)，五次方程就可解。可见 Galois 理论本身还有很多问题。

例 3 因式分解。看起来这好象是个简单问题。按非构造性的观点(或通常代数书中所写的)，只考虑把整数环推广成一种唯一因子分解环(UFD)，再研究 UFD 的一些性质(如 UFD 的多项式环仍是 UFD 等)，可是对具体的环如何进行分解就不管了。近十年来由于计算机的影响，这问题重新引起注意，提出许多方法，效率并不高。

例 4 代数几何。这是目前非常活跃的领域，有相当多的流派与研究代数几何的方法，可是除了 J. F. Ritt 的以外，基本上都是非构造性的。例如他们一开始就假定代数簇分解为不可分的或不可约的分支，然后研究不可约分支的几何性质

等等。一种简单、重要的情形是由一些多项式方程定义的代数簇：

$$\begin{cases} f_1(x_1, \cdots, x_n) = 0, \\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ f_r(x_1, \cdots, x_n) = 0。\end{cases}$$

可是对这种代数簇如何分解成不可约分支并没有引起注意，研究中给出的不可约判别法并不实用。

在西方，数学研究的一个转折点是不变式论，在此以前的数学是构造性的，而且当时要求必须如此，要证明存在就必须同时给出求法，这当然是一种束缚。不变式的研究曾是 19 世纪数学的中心课题之一，在几何研究中非常重要。例如在投影几何中，取齐次坐标(x, y)，考虑 n 个点，一般用 n 次式子$\sum a_i x^i y^{1-i} = 0$表达。由 n 个点（或 a_i）可以造出一些式子，其中有几何意义的应当在线性替换下不变或者相差一个因子，这种式子就是所谓不变式。不变式可以有很多个。例如若 I_1, I_2 是，则 $I_1^2, I_1^2 + I_2^2$ 也是，一个基本问题是能否找出有限个不变式，使得其它的都可以通过多项式表示出来。当时的研究是通过大量的计算，Gordan 彻底地解决了两个变数的情形。三个或更多个变数问题变得非常复杂、困难，不过 Hilbert 相当简洁地证明了有限基的存在性，当时的构造性观点感到难以接受，后来，Hilbert 也给出了如何求的方法*。不变式研究因为过于繁难沉寂下来，但是数学不能回避这种重要问题，如 Dieudonné 所说："不变式论象只不死鸟，屡次从灰烬中复生。"（见其《Invariant Theory, Old and New》，Academic Press, Inc. 1970）。现在又开始有了新的研

---

* 尽管 Hilbert 是公理化的代表人物，他的许多工作是构造性的。如机器证明的想法由 Hilbert 数理逻辑的工作开始具体化，对只满足从属、平行关系的纯交点型定理，他实际上给出机械化的算法，现命名为 Hilbert 机械化定理。

究。例如借助于计算机，可以期望将 19 世纪不变式的计算推进一步，不变式论也会有更广泛的应用(如 Mumford 在《Geometric Invariant Theory》中用此求解代数曲线的模空间问题)。最近，在 Bull of AMS, Vol 10(1984). *No*. 1 上有介绍 19 世纪不变式论的长篇文章，提出了许多问题，一些 19 世纪有关的书又予以重印，这也是构造性数学抬头的一个迹象。

前面提到，中国古代创造了方程术，增乘开方法等构造性的方法，它的数学基本上是构造性的。作者从事几何机器证明的研究就是在中国古代数学的启发下提出问题并想出解决办法的。继承中国的数学遗产的远景相当开阔，有待后人作长时间的努力。下面以中国剩余定理为例，说明构造性与非构造性方法的区别，特别看一看流传下来的中国的方法。

中国剩余定理：若 $m_1, \cdots, m_r$ 互素，$u_1, \cdots, u_r$ 已知，则存在唯一的 $u \in [0, m_1, \cdots, m_r)$，使

$$u \equiv u_j (\text{mod } m_j), \qquad j = 1, 2, \cdots, r。$$

非构造性证明*：唯一性没有问题，看存在性。

作映射：

$$[0, m_1, \cdots, m_r) \longrightarrow Z_{m1} \oplus \cdots \oplus Z_{mr},$$
$$v_1 \longrightarrow (v(\text{mod} m_1), \cdots, v(\text{mod} m_r))。$$

映射是单的，两边都有 $M = m_1 m_2 \cdots m_r$ 个元素，因而两边是一一对应，证得存在性。

这是个很漂亮的证明，但是 v 无从知道。

构造性证明**：先求出 N_j，使

$$N_j = \begin{cases} 1(\text{mod} m_j), \\ 0(\text{mod} m_k), j \neq k。\end{cases}$$

---

* 参见《钱宝琮科学史论文集》P536—537。

** 这两个证明引自 Knuth, Arts of Programming, vol2。

$u_1N_1+u_2N_2+\cdots+u_rN_r$ 对 M 的余数就是解。

据 Euler 定理，可取

$$N_j\equiv(m_1\cdots\hat{m}_j\cdots m_r)^{\varphi(m_i)}\equiv1(\mathrm{mod}\,m_j),$$

φ（m_i）是 Euler 函数，它等于（0，m_j）中与 m_j 互素的整数的个数，能够计算出来。

这个定理来自中国，早在《孙子算经》中已有此类问题，到秦九韶《数书九章》中成型，谓之“大衍求一术”。所谓求一，就是上面求 N_j，“大衍”隐指天意或自然规律。一次同余式问题的解法是为适应天文学家修改历法的要求而产生的。秦九韶称“圣有大衍，微寓于《易》”，将其赋于一种神秘色彩。实际上，问题来源于天文，中国古代历法不象现在公历定耶稣诞生那一年为元年，而主要是根据日月运行的客观现象制定的。古代天文学家假设在远古时代有一年的冬至节气恰恰在甲子日的上午零时，并且日月合朔也与冬至节气一个时刻。有这么一天的年度称为上元，从上元到本年经过的年数称为上元积年。一个历法有了上元积年以后，任何一年的冬至节气的日名、时刻，与任何一月平朔的日名、时刻都很容易安排出来了。在既知本年冬至的日名（平支)、时刻和十一月平朔的日名时刻的条件下，推算这一年的上元积年是一个一次同余式问题。设 a 为一回归年（从冬至到冬至）日数，R_1 为从本年冬至前的甲子日零时到冬至时的日数，b 为一朔望月（从平朔到平朔）的日数，R_2 为从十一月平朔时到冬至时的日数，那么上元积年满足下列同余式组：

$$aN\equiv R_1\ (\mathrm{mod}60)$$
$$\equiv R_2\ (\mathrm{mod}b)。$$

古代著书的方式是先给出具体的问题（问)，和解答(答)，然后指出求解方法（术)。后人还可将体会与发展（包括证明）写进去（注)。到了宋代，出现了印刷术，秦九韶还

加了“草”，以说明详细的计算过程。

秦九韶的方法：

为求 N_j，先找 k_j，使

$$k_j m_1 \cdots \hat{m}_j \cdots m_r \equiv 1 \ (\mathrm{mod}\, m_j)。$$

则取 $$N_j \equiv k_j m_1 \cdots \hat{m}_j \cdots m_r$$ 即可。

兹举《数书九章》大衍类最后一道题“余米推数”如下。

问：有米铺，诉被盗去米一般三箩，皆适满，不计细数。今左壁箩剩一合，中间箩剩一升四合，右壁箩剩一合。后获贼，系甲乙丙三名。甲称当夜摸得马杓，在左壁箩，满舀入布袋，乙称踢着木履，在中箩，舀入袋。丙称摸到漆椀，在右壁箩，舀入袋，将归食用。日久不知数。索到之日，马杓满容一升九合，木履容一升七合，漆椀容一升二合。欲知所失米数，计赃结断三盗各几何。

问题归于解三元一次同余方程组。

已知：

剩米 = $u_j = 1$，14，1（合），

器容 = $m_j = 19$，17，12（合），

求：每箩米 = $u \equiv u_j \ (\mathrm{mod}\, m_j)$。

术曰：以大衍求之。

列三器所容，为元数；	元数（= 器容）
连环求等，约为定母；（辗转相除求公因子）	定母（= m_j）
以相乘，为衍母	衍母 $M = \Pi m_k$
以定各约衍母，得衍数	衍数 $M_j = M/m_j = m_1 \cdots \hat{m}_j \cdots m_r$
各满定母，去之，得奇	奇数 = Rem（衍数/定母）= Rem（M_j/m_j）
以奇定，用大衍，求得	（上式表示 M_j 对 m_j 求余数）

乘率。	乘率 = k_j，用大衍求之
以乘衍数，得用数。	k_j * 奇数≡1mod 定母
次以剩米乘用。	用数 N_j = 乘率 * 衍数 = k_jM_j
并之，为总。	剩米 = u_j
满衍母，去之，不满，为每箩米。	总数 = $\sum u_jN_j$ = ∑剩米 * 用数
各以剩米减之，余为甲乙丙盗米。	每箩米 = Rem（总数/衍母） 盗米 = 每箩米 − 剩米
并之为共失米。	失米 = ∑盗米

这种方法可用于 m_1，…，m_r 有公因子的情形。而此时西方的方法失效。

本题具体数字如下：

元　　数	19　17　12
定母 m_j	19　17　12
衍母　$M = \Pi m_j$	19×17×12 = 3876
衍数　$M_j = M/m_j$	204　228　323
奇数　Rem(M_j/m_j)	14　7　11
乘率　(用大衍)k_j	15　5　11 (k_j * Rem(M_j/m_j) = 1modm_j) 3060　1140　3553 1　14　1
用数　$N_j = k_jM_j$ 剩米 u_j	3060 + 15960 + 3553 = 22573
总 = ∑剩米 * 用数 每箩米 = Rem(总/衍母)	Rem（22573/3876）= 3193

关键一步，如何求乘率

k * 奇数≡1mod 定母。

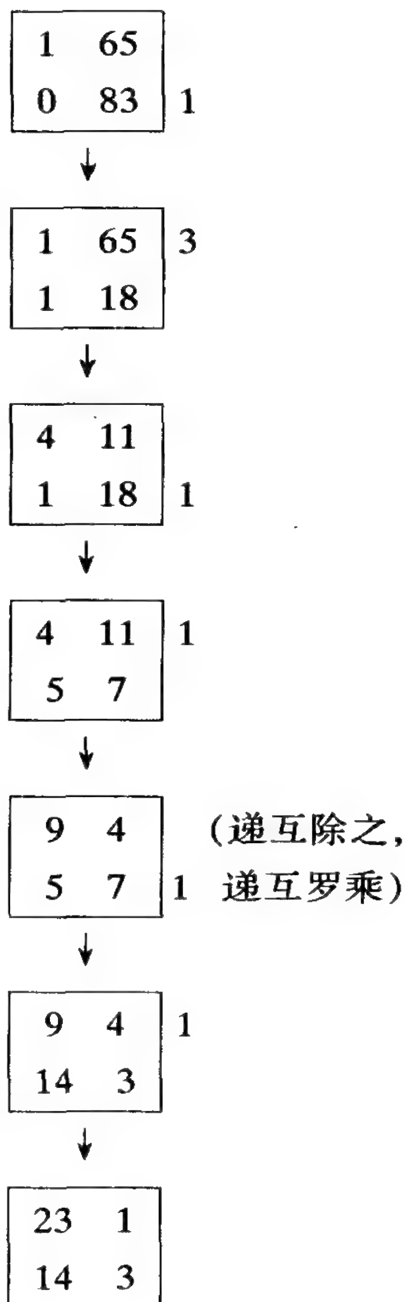
举同书第二卷中“分枭推原”为例说明。

奇数 = 65，定母 = 83。

求乘率 k，使　　$k*$奇数$\equiv 1 \bmod$ 定母。

大衍求一术云：

置奇右上，定居右下，
立天元一于左上。
所得商数，与左上相生入左下。
然后乃以右行上下，以少除多，递互除之。
所得商数，随即递互罗乘，
归左行上下。
须使右上末后奇一而止，
乃验左上所得，以为乘率，
或奇数已见单一者，
便为乘率。



乘率 $k = 23$

西法须求 $N=65^{\varphi(83)}=65^{82}$，简直无法计算，用大衍求一术做则效率相当高，其原理是：

记号 $\begin{array}{|cc|} c & r \\ c' & r' \end{array}$ 表示 $c*$奇数$\equiv r\bmod$定母，

$c'*$奇数$\equiv -r'\bmod$定母，

则

$$①\ \begin{array}{|cc|} c & r \\ c' & r' \end{array} \xrightarrow[r=qr'+r'']{r>r'} \begin{array}{|cc|} c+qc' & r'' \\ c' & r' \end{array}$$

证 $r+q(-r')=r''\Rightarrow(qc'+c)*$奇$\equiv r''\bmod$定母。

$$2\ \begin{array}{|cc|} c & r \\ c' & r' \end{array} \xrightarrow[r='qr+r'']{r<r'} \begin{array}{|cc|} c & r \\ c'+qc & r'' \end{array}$$

证　$qc+(-r')=-r''\Rightarrow(qc+c')*$奇$\equiv -r''\bmod$定母。

（本文是胡森、王东明根据作者在中科院研究生讲授《机器证明》课中一节整理而成，已经本人审阅。）

分角线相等的三角形*

——初等几何机器证明问题

一、一个古老问题："两条内分角线相等的三角形是等腰三角形"

初中课本里有一道题目：

求证等腰三角形两个底角的内分角线** 相等。

这是初中生都会解决的简单问题。

但是，这个命题的逆命题：

两条内分角线相等的三角形是等腰三角形，

看起来十分简单，却并不是容易证明的。如果不信，请你自己试一试。

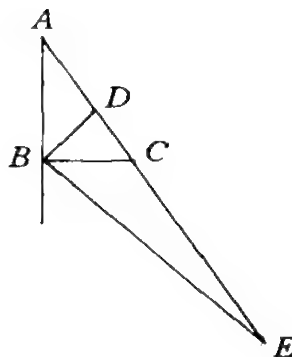
这是一个古老问题。

从古至今，有很多人用了很多方法证明这个逆命题，但都

---

* 本文摘自《分角线相等的三角形》（吴文俊、吕学礼著），人民教育出版社，1985年。

** 三角形一个角的内分角线是指这个内角的平分线从顶点到与对边的交点间的线段，外分角线是指外角的平分线从顶点到与对边的交点间的线段。如图的$\triangle ABC$中，BD是一条内分角线，BE是一条外分角线。



不是轻而易举的。下面举出其中比较起来简单些的几种证法。

1．利用反证法

已知在$\triangle ABC$中，内分角线BD和CE相等。求证$AB=AC$（如图1)。

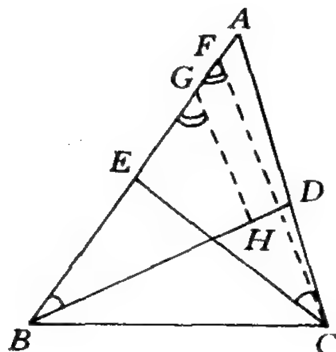


图1

证明：假定$AB\neq AC$。且设$AB>AC$。则$\angle ACB>\angle ABC$。因而$\frac{1}{2}\angle ACB>\frac{1}{2}\angle ABC$，即$\angle ACE>\angle ABD$。

在$\angle ACE$内，作$\angle ECF=\angle DBA$。则因$\angle ECB>\angle DBC$（分别等于$\frac{1}{2}\angle ACB$和$\frac{1}{2}\angle ABC$)，所以仍有$\angle FCB>\angle FBC$。故得$BF>CF$。

在BF上取G点，使$BG=CF$。再作$GH\parallel FC$，与BD交于H。

那么在$\triangle BGH$和$\triangle CFE$中，$\angle GBH=\angle FCE$，$BG=CF$、$\angle BGH=\angle CFE$，因此$\triangle BGH\cong\triangle CFE$，而得$BH=CE$。

这和$BD=CE$相矛盾。

因此不可能有$AB>AC$。

同样，如果假定$AB<AC$，那么也会产生矛盾，因此也不可能有$AB<AC$。

这样就证明了$AB=AC$。

2．利用直接证法

已知在$\triangle ABC$中，内分角线BD和CE相等。求证$AB=AC$。

证明：命$\angle ABC=2\alpha$，$\angle ACB=2\beta$。

作 DF 使 $\angle BDF=\beta$，作 BF 使 $\angle DBF=\angle CEB$，如图 2。

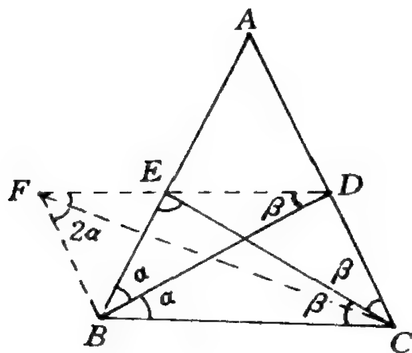


图 2

设 DF 与 BF 交于 F。则在 $\triangle DBF$ 和 $\triangle CEB$ 中，$\angle BDF=\angle ECB$，$\angle DBF=\angle CEB$，且 $BD=EC$，故 $\triangle DBF\cong\triangle CEB$，因此，$\angle BFD=\angle EBC=2\alpha$，$BF=EB$。

$$\begin{aligned}\angle FBC&=\angle FBD+\alpha=\angle BEC+\alpha=180^\circ-(2\alpha+\beta)+\alpha\\&=180^\circ-(\alpha+\beta),\end{aligned}$$

$\angle CDF=\angle CDB+\beta=180^\circ-(\alpha+2\beta)+\beta=180^\circ-(\alpha+\beta)$，并且 $2\alpha+2\beta<180^\circ$，$\therefore \alpha+\beta<90^\circ$，$\therefore \angle FBC=\angle CDF>90^\circ$。又 $BC=DF$，$FC=CF$，$\therefore \triangle BCF\cong\triangle DFC$（两个三角形有两边及其中一边的对角分别相等，且这个对角为钝角，则两三角形全等）。$\therefore BF=DC$。

又因 $BF=EB$，所以 $DC=EB$。又 $DB=EC$，$BC=CB$，因此 $\triangle DBC\cong\triangle ECB$。

$\therefore \angle DCB=\angle EBC$，$\therefore AB=AC$。

3．代数证法

已知、求证同(1)、(2)。设 BC、CA、AB 的长分别为 a、b、c。则由内分角线的长的公式，根据 $BD=CE$，可得

$$\frac{2}{a+c}\sqrt{acs(s-b)}=\frac{2}{a+b}\sqrt{abs(s-c)}, \tag{1}$$

其中　$s=\frac{1}{2}(a+b+c)$。

由(1)可得$(a+b)^2c(s-b)=(a+c)^2b(s-c)$,

化简,$a^2cs+b^2cs-2ab^2c-b^3c=a^2bs+c^2bs-2abc^2-c^3b$,

即　$a^2s(c-b)+bcs(b-c)-2abc(b-c)-bc(b^2-c^2)=0$,

$(b-c)(a^2s-bcs+2abc+b^2c+bc^2)=0$,

$(b-c)(a^2s-bcs+abc+bc\cdot 2s)=0$,

$\therefore\quad (b-c)(a^2s+abc+bcs)=0$。　　(2)

因 $a>0,b>0,c>0,s>0$,故(2)式左边第二个因式 $a^2s+abc+bcs>0$。

于是由(2)式可得 $b-c=0$,即$\triangle ABC$ 为等腰三角形。

二、传统证法与机器证法

上节我们提到了一个命题"等腰三角形两个底角的内分角线相等"和它的逆命题"两条内分角线相等的三角形是等腰三角形"。原命题容易证明,而这个逆命题很难证明,这个逆命题成为一个古老问题,很多人想了很多方法加以证明。

读者们也许都有这个经验,用传统方法证明每一个稍难的初等几何命题,都要经过一番巧思。有的要添这样的辅助线,有的要添那样的辅助线。不同的命题有不同的证法,也就需要不同的巧思。一个新的命题需要通过新的巧思,找到新的证法。

下面我们粗略介绍利用电子计算机来证明初等几何命题的方法,简称机器证法。这种方法不是特殊地适用于个别的命题,而是普遍地适用于初等几何的所有命题,至少是某一类的很多命题。只要按照这种方法机械地进行,在有限步之后,就可对这一类中的任何初等几何命题判定它是真是假。命题是真的,这

个命题就得到了证明;命题是假的,这个命题就得到了否定。机器证明只需机械地进行,对于这类中的任何命题都是按照同样的步骤进行,不必对特殊的命题运用特殊的巧思。

初等几何的传统证明与机器证明之间的关系,就象应用题的算术解法与代数解法之间的关系。

为了通俗易懂,本文只能对机器证明介绍一些大致情况,读者如要了解它的详细情况以及基本原理,可请进一步参阅《几何定理机器证明的基本原理(初等几何部分)》(科学出版社出版,吴文俊著)等书。

三、机器证法举例

为了说明机器证法的大意,我们先举一个极简单的例子。这个例子用传统证法完全可以轻易地解决。以它为例,只是为了说明机器证法的大致情况,为下面进一步说明机器证法作些准备。

求证:平行四边形的对角线互相平分。

设 $ABCD$ 为平行四边形,求证 AC、BD 互相平分(如图 3)。

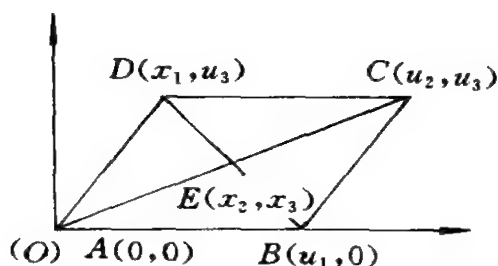


图 3

机器证法与解析几何方法相仿,也是先把几何问题化为代数问题。

以 A 点为原点 O,以 AB 为横轴,作成直角坐标系。

则 A 点的坐标为$(0,0)$。设 B 点的坐标为$(u_1,0)$，C 点的坐标为(u_2,u_3)，D 点的坐标为(x_1,u_3)，对角线 BD 与 AC 的交点 E 的坐标为(x_2,x_3)。

这里 u_1,u_2,u_3 是独立坐标，而 x_1,x_2,x_3 是随着 u_1,u_2,u_3 而确定的，所以不是独立坐标。D 的纵坐标取得和 C 的纵坐标一样，都是 u_3，这是已经运用了 $DC /\!/ AB$ 的条件。在这里，这样做是为了使后面的说明更为简化而采取的特殊处理。实际上完全可以不必运用 $DC /\!/ AB$ 的条件，而把 D 的坐标取作(x_1,x_2)，E 的坐标取作(x_3,x_4)，后面的说明仍能完全适用，不过略繁一些。

取好坐标以后，我们先把本题的题设部分表示成代数形式。

$AD /\!/ BC$ 可以用$\dfrac{u_3-0}{x_1-0}=\dfrac{u_3-0}{u_2-u_1}$表示(斜率相等)，但在机器证明中，不准许用分式，而把 $AD /\!/ BC$ 用关于整式的等式来表示：

$$(u_3-0)(u_2-u_1)-(x_1-0)(u_3-0)=0,$$

即

$$u_3(u_2-u_1)-x_1u_3=0。\tag{1}$$

要注意等式(1)实际上并不只是表示一般的几何意义下的 $AD /\!/ BC$，而是表示 $AD /\!/ BC$，或 A 与 D 重合，或 B 与 C 重合，或 AD 与 BC 在同一条直线上等几种情况。这样做，虽然好象是扩大了我们原来关于平行的概念，但却有利于研究更为普遍的情况，并且也有利于找到关于机器证明的一个一般步骤。

E 点在 BD 上，在机器证明中可以用下面关于整式的等式来表示：

$$x_3(x_1-u_1)-u_3(x_2-u_1)=0。\tag{2}$$

同样，E 点在 AC 上，可以用下式表示：

$$x_3u_2-x_2u_3=0。\tag{3}$$

(1)、(2)、(3)三式表示本题的题设部分(题设 $DC /\!/ AB$ 已在取坐标时用过了)。

下面再把本题的终结部分表示成代数形式。

$EA = EC$ 就是(机器证明中不用根式,改用 $EA^2 = EC^2$,下同)

$$x_2^2 + x_3^2 = (u_2 - x_2)^2 + (u_3 - x_3)^2,$$

即
$$u_2^2 - 2u_2x_2 + u_3^2 - 2u_3x_3 = 0。\qquad (4)$$

$EB = ED$ 就是

$$(x_2 - u_1)^2 + x_3^2 = (x_1 - x_2)^2 + (u_3 - x_3)^2,$$

即
$$-2u_1x_2 + u_1^2 - x_1^2 + 2x_1x_2 - u_3^2 + 2u_3x_3 = 0。\qquad (5)$$

现在的问题就是怎样从本题的题设部分(1)、(2)、(3)推出本题的终结部分(4)、(5)。

我们通常的做法是,从(1)、(2)、(3)中解出 x_1、x_2、x_3,得到,在 $u_1 \neq 0$, $u_3 \neq 0$ 的条件下*,

$$\begin{cases} x_1 = u_2 - u_1, \\ x_2 = u_2/2, \\ x_3 = u_3/2。 \end{cases}$$

把它们代入(4)式的左边,得

$$u_2^2 - 2u_2 \cdot \frac{u_2}{2} + u_3^2 - 2u_3 \cdot \frac{u_3}{2},$$

这个式子恒等于 0,所以(4)式成立。

把它们代入(5)式的左边,得

$$-2u_1 \cdot \frac{u_2}{2} + u_1^2 - (u_2 - u_1)^2 + 2(u_2 - u_1) \cdot \frac{u_2}{2} - u_3^2 + 2u_3 \cdot \frac{u_3}{2},$$

---

* 从图 3 可以看到,$u_1 \neq 0$ 的几何意义是 B 点与 A 点不相重合;$u_3 \neq 0$ 的几何意义是 DC 与 AB 不在同一条直线上。对于我们所理解的平行四边形来说,这些条件显然是符合的。

这个式子也恒等于 0,所以(5)式也成立。

这就证明了在 $u_1 \neq 0$, $u_3 \neq 0$ 的条件下,本题中的命题是真的。

上面所说的是通常的做法,还不是机器证明。

机器证明用的是下面的方法。虽然对有些问题机器证明的方法似乎反而比通常的做法更繁,但是因为机器证明的方法有一定的步骤,可以通过一定的程序让机器来运行,而且对于较为复杂的问题通常的做法很难进行甚至无法进行,而机器证法则不管问题是简单的还是复杂的,都能同样按照一定的步骤进行,所以机器证明所用的下面的方法可以解决通常的做法不易解决或无法解决的问题。

方法如下。

第一步　先对命题的题设部分(1)、(2)、(3)三式,即

$$u_3(u_2 - u_1) - x_1 u_3 = 0, \tag{1}$$

$$x_3(x_1 - u_1) - u_3(x_2 - u_1) = 0, \tag{2}$$

$$x_3 u_2 - x_2 u_3 = 0 \tag{3}$$

进行整理,使其中一个式子只能包含一个 x,例如 x_1,一个式子只能包含这个 x 和另一个 x,例如 x_1 和 x_2(或 x_1 和 x_3),而一个式子可以包含三个 x,即 x_1, x_2, x_3。

对于这里的(1),(2),(3),这是很容易做到的。

(1)式中本来只含 x_1。在 $u_3 \neq 0$ 的条件下,(1)式就是

$$u_2 - u_1 - x_1 = 0。$$

我们把它的左边(右边是零,下同)叫做 f_1,那么(1)式就是(在 $u_3 \neq 0$ 的条件下)

$$f_1 = u_2 - u_1 - x_1 = 0。\tag{1'}$$

(2)、(3)两式都是既有 x_2,又有 x_3。我们可以从中消去 x_2。由(2)式减去(3)式,即得

$$x_3(x_1-u_1-u_2)+u_1u_3=0。$$

我们把它的左边叫做 f_2，就得

$$f_2=x_3(x_1-u_1-u_2)+u_1u_3=0。\qquad (2')$$

再把(3)式的左边叫做 f_3，就得

$$f_3=x_3u_2-x_2u_3=0。\qquad (3')$$

上面的(1′)只含 x_1，(2′)只含 x_1 和 x_3，(3′)含 x_1，x_2 和 x_3(实际上不含 x_1)。这就达到了整理的目的。这样的整理叫做**三角化***。意思就是化成

$$\begin{matrix} x_1 \\ x_1, x_3 \\ x_1, x_2, x_3 \end{matrix} \qquad \left(\text{或} \quad \begin{matrix} x_1 \\ x_1, x_2 \\ x_1, x_2, x_3 \end{matrix}\right)$$

的形式。

把题设部分三角化以后，得到三个多项式：

$$f_1=u_2-u_1-x_1,$$
$$f_2=x_3(x_1-u_1-u_2)+u_1u_3,$$
$$f_3=x_3u_2-x_2u_3。$$

然后进行第二步。

第二步　把表示题目的终结部分的等式的左边(右边是 0)用 g 表示。例如把(4)式的左边用 g_1 表示，即令

$$g_1=u_2^2-2u_2x_2+u_3^2-2u_3x_3。$$

((5)式的左边用 g_2 表示，以后再说。)

再把 g_1 除以 f_3(两者都看作 x_2 的多项式)，所得的余式除以 f_2(两者都看作 x_3 的多项式)，又一次所得的余式除以 f_1(两者都看作 x_1 的多项式)，最后把所得的余式叫做 R，看 R 是不是恒等于零。

---

* 严格说来，三角化还有其他一些要求，这里从略。

实际进行如下。

把 g_1 除以 f_3(都看作 x_2 的多项式)。为了避免分式,我们用 u_3g_1 除以 f_3,即

$$
\begin{array}{r|l}
 & 2u_2 \\
\hline
-u_3x_2+u_2x_3u & -2u_2u_3x_2-2u_3^2x_3+u_2^2u_3+u_3^3 \\
 & -2u_2u_3x_2+2u_2^2x_3 \\
\hline
 & -2u_3^2x_3-2u_2^2x_3+u_2^2u_3+u_3^3
\end{array}
$$

这就得到

$$u_3g_1=2u_2f_3+R_3, \tag{6}$$

其中 $R_3=-2(u_2^2+u_3^2)x_3+u_3(u_2^2+u_3^2)$。

把 R_3 除以 f_2(都看作 x_3 的多项式),得到(除法算式从略,下同)

$$(x_1-u_1-u_2)R_3=-2(u_2^2+u_3^2)f_2+R_2, \tag{7}$$

其中 $R_2=2u_1u_3(u_2^2+u_3^2)+u_3(u_2^2+u_3^2)(x_1-u_1-u_2)$。

再把 R_2 除以 f_1(都看作 x_1 的多项式),得到

$$R_2=-u_3(u_2^2+u_3^2)f_1+R, \tag{8}$$

而其中的 R 恒等于 0!

这就说明了在某些附加条件下,从本题的题设部分(1)、(2)、(3)三式可以推出本题的终结部分(4)式。

这是因为由(6)式可得

$$u_3(x_1-u_1-u_2)g_1=2u_2(x_1-u_1-u_2)f_3+(x_1-u_1-u_2)R_3,$$

以(7)式代入,得

$$u_3(x_1-u_1-u_2)g_1=2u_2(x_1-u_1-u_2)f_3-2(u_2^2+u_3^2)f_2+R_2,$$

再以(8)式代入,得

$$u_3(x_1-u_1-u_2)g_1=2u_2(x_1-u_1-u_2)f_3-2(u_2^2+u_3^2)f_2$$

$$-u_3(u_2^2+u_3^2)f_1+R。\quad (9)$$

(9)式可以写成下列形式：

$$c_1c_2g_1=a_1f_3+a_2f_2+a_3f_3+R。\quad (10)$$

其中 $c_1=u_3$, $c_2=x_1-u_1-u_2$, $a_1=2u_1(x_1-u_1-u_2)$, $a_2=-2(u_2^2+u_3^2)$, $a_3=-u_3(u_2^2+u_3^2)$。

由命题的题设部分(1)、(2)、(3)三式，可以推得 $f_1=0$, $f_2=0$, $f_3=0$；而由这里的除法计算，又得 $R=0$。因此，由(10)式可得

$$c_1c_2g_1=a_1\cdot 0+a_2\cdot 0+a_3\cdot 0+0,$$

即

$$c_1c_2g_1=0。$$

由此可知，只要 $c_1\neq 0$, $c_2\neq 0$, 就可推得本题的终结部分(4)式：

$$g_1=0。$$

这里的 $c_1\neq 0$, $c_2\neq 0$ 叫做附加条件。

上面所说的第二步，叫做**逐次除法**。本题第二步逐次除法的结果，得到恒等于零的余式 R。这就可以断定，在某些附加条件下，由本题的题设部分(1)、(2)、(3)三式可以推出本题的终结部分(4)式。

$c_1\neq 0$ 就是 $u_3\neq 0$, 它的几何意义就是 DC 和 AB 不在同一直线上(如图 3)。$c_2\neq 0$ 就是 $x_1-u_1-u_2\neq 0$, 而由 $f_1=0$ 可得 $x_1=u_2-u_1$, 因此 $c_2\neq 0$ 就等价于 $u_1\neq 0$, 它的几何意义就是 B 点不与 A 点重合(图)。因此，对于我们一般理解的平行四边形来说，这两个附加条件 $c_1\neq 0$ 和 $c_2\neq 0$ 是成立的，由此本题的终结部分(4)式得证。((5)式的证明见后)

从上面的例子可以看到，机器证明的思路是：先把几何问题化为代数问题，列出表示命题的题设部分和终结部分的只含整式的等式。这一步可由人工进行，一般不很烦难(也可编成程

序，由机器进行）。然后进行三角化和逐次相除，这两步可以由机器按一定的程序进行。虽然我们用人工进行这两步觉得较繁，但是机器进行却是不怕繁的。即使余式多达几十项、几百项，它都不怕。如果最后得到的余式恒等于零，那就说明在某些附加条件下，从命题的题设部分可以推出终结部分，也就是说，在某些附加条件下，命题已经得到证明。

四、机器证法大意

上节我们举了机器证明初等几何命题的一个例子。本节借助这个例子来说明机器证明的大意。

通过上节的例子可以看到，机器证明初等几何命题，首先要把几何问题化成代数形式，用关于整式的等式写出命题的题设部分和终结部分。这是可由人工或机器进行的。

适当选择坐标系，即选取图形中适当的点作为原点，适当的直线作为坐标轴，往往可使后面的各个步骤以及问题的解决得到很大的简化。

选定坐标系后，我们可以把独立的各点（即它们的位置是可以任意选取的）的坐标用（u_1，u_2），（u_3，u_4）等表示。而把不独立的各点（即它们的位置是由其他点的位置所确定的）的坐标用（x_1，x_2），（x_3，x_4）等表示。哪些点是独立点，哪些点不是独立点，往往不是固定的。如果取某些点为独立点，那么其他点就不是独立点；如果取另一些点为独立点，那么其他的点就不是独立点。但在同一个问题中，独立坐标的个数则是确定的。当然，独立点的坐标和不独立点的坐标并不一定要用 u 和 x 来区分，但是这样区分一下，可以使后面的说明得到方便。

确定了点的坐标以后，初等几何中的一些关系（无论在题

设中，或在终结中）可以化成代数形式如下。

1．线段相等

如图 4,线段 AB 的长度(即 A、B 两点的距离)的平方为

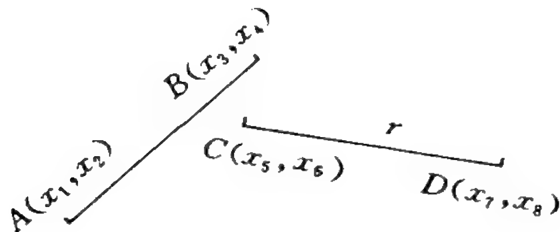


图 4

$$(x_1-x_3)^2+(x_2-x_4)^2。$$

线段 CD 的长度为 r,可以表示为

$$(x_5-x_7)^2+(x_6-x_8)^2-r^2=0。$$

线段 AB 与 CD 相等,可以改用它们的长度的平方相等表示为

$$(x_1-x_3)^2+(x_2-x_4)^2-(x_5-x_7)^2-(x_6-x_8)^2=0。$$

2．一点是两点的中点

如图 5,B 是 A、C 的中点,可以表示为

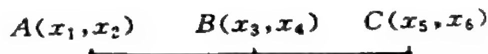


图 5

$$\begin{cases}2x_3-x_1-x_5=0,\\2x_4-x_2-x_6=0。\end{cases}$$

上面的两个等式在一起实际上表示“B 是 A、C 的中点”或“A、B、C 三点重合”。

3．两线平行

如图 6,$AB /\!/ CD$ 可以表示为

$$(x_4-x_2)(x_7-x_5)-(x_3-x_1)(x_8-x_6)=0。$$

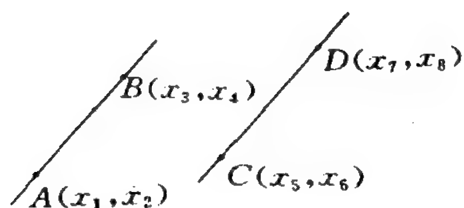


图 6

这个等式可从 AB 与 CD 的斜率相等导出，但这个等式所表示的实际上是："$AB // CD$"，或"A 与 B 重合"，或"C 与 D 重合"，或"AB 与 CD 在同一条直线上"。

4. 两线垂直

如图 7，$AB \perp CD$ 可以表示为

$$(x_4-x_2)(x_8-x_6)+(x_3-x_1)(x_7-x_5)=0。$$

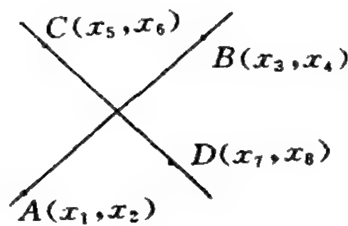


图 7

这个等式可从 AB 与 CD 的斜率互为负倒数导出，但这个等式所表示的实际上是："$AB \perp CD$"，或"A 与 B 重合"，或"C 与 D 重合"。

5. 三点共线

如图 8，A、B、C 三点共线可以表示为

$$(x_4-x_6)(x_1-x_5)-(x_3-x_5)(x_2-x_6)=0。$$

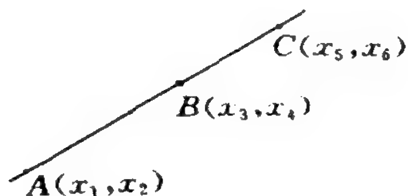


图 8

这个等式可以从 BC 与 AC 的斜率相等导出，但这个等式所表示的实际上是："A、B、C 三点共线"，或"其中有两点重合"，或"三点重合"。

熟悉行列式的读者可以看到，上面的等式与下面关于行列式的等式是一样的：

$$\begin{vmatrix} x_1 & x_2 & 1 \\ x_3 & x_4 & 1 \\ x_5 & x_6 & 1 \end{vmatrix} = 0。$$

6．三线共点

如图 9，AB、CD、EF 三条直线都经过 G 点，可以用三个等式来表示，这三个等式分别表示 A、B、G 三点共线，C、D、G 三点共线，以及 E、F、G 三点共线。

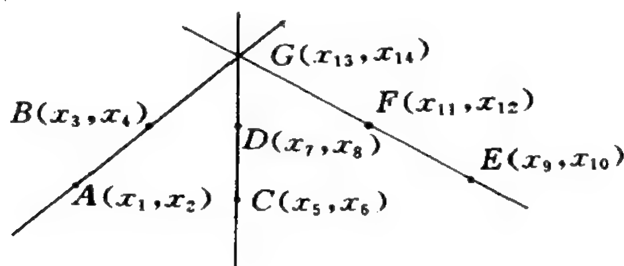


图 9

7．两角相等

如图 10，两角 ABC 和 DEF 相等，不但需要它们的度数相等，而且需要它们从始边（BA、ED）旋转到终边（BC，EF）的方向也相同（都是逆时针方向，或都是顺时针方向）。

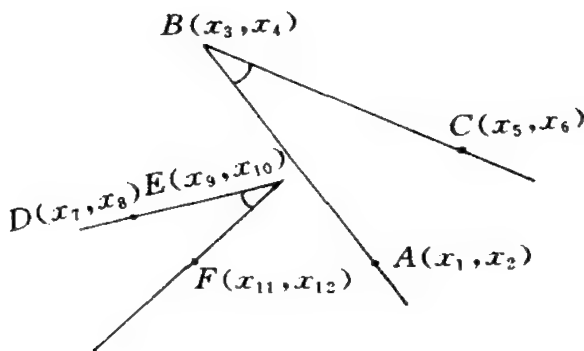


图 10

如果一个角从始边旋转到终边的方向是逆时针方向，那么

这个角的正切值是$\frac{S_{终}-S_{始}}{1+S_{终}S_{始}}$，这里 $S_{始}$、$S_{终}$ 分别表示这个角的始边和终边的斜率。

按此，图 10 中的$\angle ABC=\angle DEF$（$\angle ABC$ 的始边为 BA，终边为 BC，其他同）可以表示为

$[(x_6-x_4)(x_1-x_3)-(x_2-x_4)(x_5-x_3)][(x_{11}-x_9)(x_7-x_9)+(x_{12}-x_{10})(x_8-x_{10})]-[(x_{12}-x_{10})(x_7-x_9)-(x_8-x_{10})(x_{11}-x_9)][(x_5-x_3)(x_1-x_3)+(x_6-x_4)(x_2-x_4)]=0$。

这个等式所表示的实际上是：

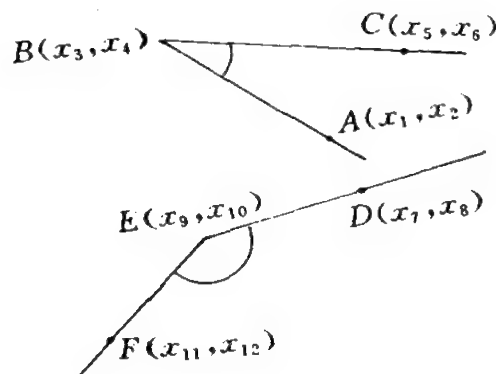


图 11

"A 与 B 重合"，或"C 与 B 重合"，或"D 与 E 重合"，或"F 与 E 重合"，或"$\angle ABC$ 与$\angle DEF$ 度数相同且旋转方向相同"（如图 10），或"$\angle ABC$ 与$\angle DEF$ 度数相补且旋转方向相反"（如图 11）。

8. 点在分角线上

如图 12，P 点在$\angle ABC$ 的分角线上，可以表示为

$[(x_8-x_4)(x_1-x_3)-(x_2-x_4)(x_7-x_3)][(x_5-x_3)(x_7-x_3)+(x_6-x_4)(x_8-x_4)]-[(x_6-x_4)(x_7-x_3)-(x_8-x_4)(x_5-x_3)][(x_7-x_3)(x_1-x_3)+(x_8-x_4)(x_2-x_4)]=0$。

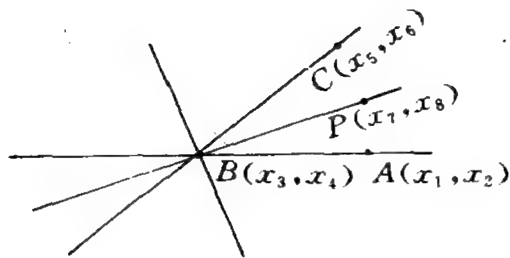


图 12

这个等式可从考虑∠*ABP* 的正切与∠*PBC* 的正切相等而得出。它所表示的实际上是“*A* 与 *B* 重合”，或“*C* 与 *B* 重合”，或“*A* 与 *C* 重合”，或“*P* 在∠*ABC* 的分角线上”，或“*P* 在∠*ABC* 的邻补角的分角线上”。

9．线段的比相等

如图 13，线段 *AB* 与 *CD* 的(长度的)比等于 r，可以表示为

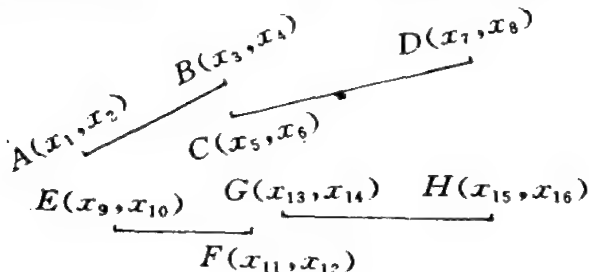


图 13

$$(x_1-x_3)^2+(x_2-x_4)^2-r^2[(x_5-x_7)^2+(x_6-x_8)^2]=0。$$

这个等式实际上也表示“*A* 与 *B* 重合且 *C* 与 *D* 重合”的情形。

又线段 *AB* 与 *CD* 的比等于线段 *EF* 与 *GH* 的比(即线段 *AB*，*CD*，*EF*，*GH* 成比例)，可以表示为

$$[(x_1-x_3)^2+(x_2-x_4)^2][(x_{13}-x_{15})^2+(x_{14}-x_{16})^2]$$
$$-[(x_5-x_7)^2+(x_6-x_8)^2][(x_9-x_{11})^2+(x_{10}-x_{12})^2]=0。$$

这个等式就是表示线段 *AB* 与 *GH* 的积等于线段 *CD* 与 *EF* 的积，实际上也表示“*A* 与 *B* 重合且 *C* 与 *D* 重合”，或“*A* 与 *B* 重合且 *E* 与 *F* 重合”，或“*G* 与 *H* 重合且 *C* 与 *D* 重合”，或“*G* 与 *H*

重合且 E 与 F 重合”等情形。

10．一点分两点成定比

如图 14，P 点分 A、B 两点成定比 r，是指 P 在 A、B 所在的直线上，且有向线段 AP 与有向线段 PB 的比是 r。因此，P 点在 A、B 之间时 r 为正(如图 14 的上图)，P 点在 A、B 之外时 r 为负(如图 14 的下图)。

P 点分 A、B 成定比 r，可以表示为

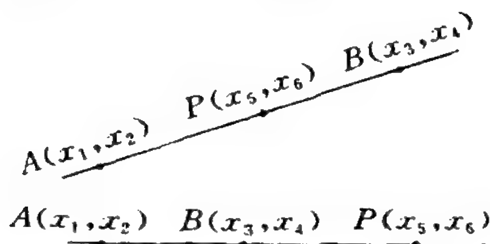


图 14

$$\begin{cases} x_5 - x_1 - r(x_3 - x_5) = 0, \\ x_6 - x_2 - r(x_4 - x_6) = 0。 \end{cases}$$

这两个等式在一起实际上表示：“P 分 A、B 成定比 r”，或“A、B、P 三点重合”，或“A 与 B 重合且 P 不与 A、B 重合”。

11．一点在一圆上

如图 15，P 点在以 O 为圆心，以 r 为半径的圆上，可以表示为

$$(x_3 - x_1)^2 + (x_4 - x_2)^2 = r^2。$$

当 r 为零时，这个等式表示 P 与 O 重合。

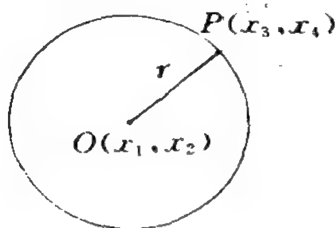


图 15

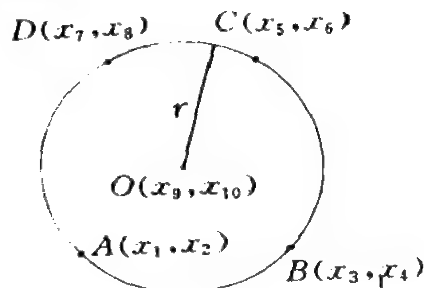


图 16

12. 四点共圆

如图 16，四点 A、B、C、D 共圆，可以设这个圆的圆心为 O，半径为 r，而把这四点共圆表示为

$$\begin{cases}(x_1-x_9)^2+(x_2-x_{10})^2=r^2,\\(x_3-x_9)^2+(x_4-x_{10})^2=r^2,\\(x_5-x_9)^2+(x_6-x_{10})^2=r^2,\\(x_7-x_9)^2+(x_8-x_{10})^2=r^2。\end{cases}$$

也可利用角的旋转方向及相等相补关系，把这四点共圆用本节"7. 两角相等"中表示两角相等的式子来表示。例如在图 16 中 A、B、C、D 四点这样的位置情况下，$\angle ABD$ 与 $\angle ACD$ 旋转方向相同且度数相等，$\angle ABC$ 与 $\angle ADC$ 旋转方向相反且度数相补。

上面我们举出了把几何关系用代数形式来表示的几种情况。这里已经基本上包括了通常遇到的一些几何关系。如果另外遇到这里没有提及的几何关系时，也可仿照这里的方法，把它们化成代数形式。

当要证明一个命题时，可以先把这个命题的题设部分和终结部分分别化成代数形式。这个工作可由人工或机器来进行。然后可以利用机器按照一定的程序进行下面两步工作。

第一步：三角化

把表示命题的题设部分的式子进行整理，得到一组关于整式的等式：

$$\begin{gathered}f_1=0,\\f_2=0,\\f_3=0,\\\cdots\cdots\\f_n=0。\end{gathered}$$

其中各个 $f_i(i=1,2,3,\cdots,n)$都可以含有独立坐标 $u_1,u_2,\cdots$，但 f_1 只含有一个不独立坐标 x_1，f_2 只含有 x_1 和另一个不独立坐标 x_2，f_3 只含有 x_1，x_2 和另一个不独立坐标 $x_3,\cdots,f_n$ 可以含有所有的不独立坐标(设有 n 个)。当然上面所说的不独立坐标不一定按 $x_1,x_2,x_3,\cdots,x_n$ 的排列次序，而是可以按任何其他的排列次序。

第二步：逐次除法

把表示命题的终结部分的式子写成

$$g=0。$$

把 g 除以 f_n(都看作 x_n 的多项式)，得

$$c_1g=a_1'f_n+R_n;$$

再把 R_n 除以 f_{n-1}(都看作 x_{n-1}的多项式)，得

$$c_2R_n=a_2'f_{n-1}+R_{n-1};$$

再把 R_{n-1}除以 f_{n-2}(都看作 x_{n-2}的多项式)，得

$$c_3R_{n-1}=a_3'f_{n-2}+R_{n-2};$$

$$\cdots$$

同样进行，最后得

$$c_{n-1}R_2=a'_{n-1}f_2+R_2;$$

$$c_nR_3=a_n'f_1+R。$$

把上面的第一式乘以 c_2，再把第二式代入，得

$$c_1c_2g=a_1'c_2f_n+a_2'f_{n-1}+R_{n-1};$$

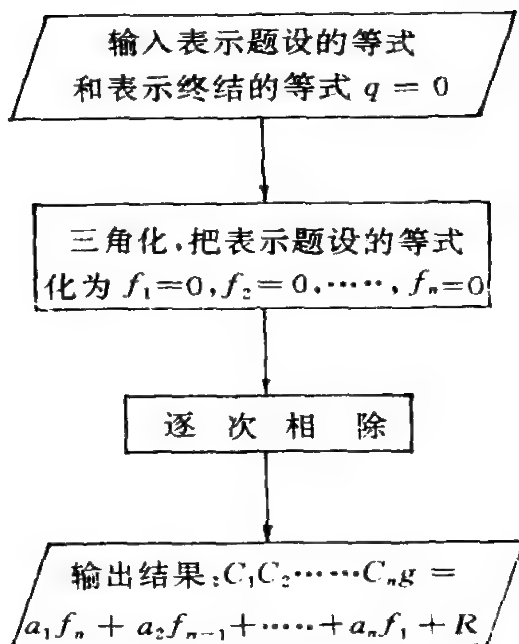
再把这个式子乘以 c_2，然后把第三式代入，得

$$c_1c_2c_3g=a_1'c_2c_3f_n+a_2'c_3f_{n-1}+a_3'f_{n-2}+R_{n-2};$$

依此进行，最后得

$$c_1c_2c_3\cdots c_ng=a_1f_n+a_2f_{n-1}+\cdots+a_nf_1+R。\tag{1}$$

上面的第一步三角化和第二步逐次除法是可以通过计算机来完成的。就是说，



这个过程，可以由计算机来完成，而且是切实可行的。就是说，不必动用大型的或超级的计算机，不必经年累月，耗费大量时间，即使对于一般认为是初等几何中的难题，只要利用普通的微型计算机，在合理的时间（例如几小时）以内就可完成。

从输出的结果(1)式，可以看到，如果最后所得的余式 R 恒等于 0，那么在题设条件 $f_1=0, f_2=0, \cdots, f_n=0$ 以及附加条件 $c_1\neq 0, c_2\neq 0, \cdots, c_n\neq 0$ 下，一定可以得到终结条件 $g=0$。这就是说，在 $c_1\neq 0, c_2\neq 0, \cdots, c_n\neq 0$ 的附加条件下，命题已经得到证明。反之，如果没有得出余式 R 恒等于 0，那么命题的成立与否还不一定，就是说，命题还在存疑之中。

以上只是介绍了机器证法的一个梗概。就是说，利用传统方法需要冥思苦想才能得到证明的一些初等几何命题，可以利用机器，通过一定的程序加以证明。

五、机器证法再举例

上节通过第三节的举例，说明了机器证法的大意。现在再举一些例子如下。下面一些例子，有的实际上根本用不到机器，只用人工就能解决，但是所用的方法则是机器证明所用的方法。有的虽要借助于一些机器计算，但也是比较简单的。

先举一个十分简单的例子。

例 1　求证：三角形的三条高线共点。

如图 17，选取坐标系，以$\triangle ABC$的AB边所在的直线为横轴，过C点的高线为纵轴。设过A点的高线与过C点的高线相交于H，且设各点的坐标为$A(u_1,0)$，$B(u_2,0)$，$C(0,u_3)$，$H(0,x_1)$。

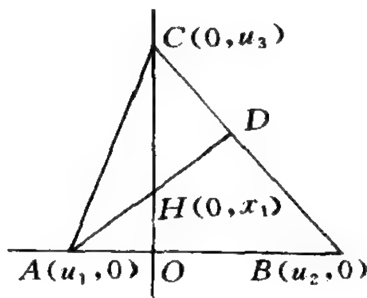


图 17

只要证明$BH\perp AC$，本题即可得证。

由题设$AH\perp BC$，得

$$f_1=(x_1-0)(u_3-0)+(0-u_1)(0-u_2)=0。$$

而终结$BH\perp AC$，即

$$g=(x_1-0)(u_3-0)+(0-u_2)(0-u_1)=0。$$

可见f_1与g完全相同。

因此，g除以f_1所得的余式R当然为 0，本题中的命题已被证明。

例 2　证明第三节中的(5)式。

这就是，从

$$f_1=u_2-u_1-x_1=0,$$

$$f_2=x_3(x_1-u_1-u_2)+u_1u_3=0,$$

$$f_3 = x_3 u_2 - x_2 u_3 = 0$$

推出

$$g_2 = -2u_1x_2 + u_1^2 - x_1^2 + 2x_1x_2 - u_3^2 + 2u_3x_3 = 0。$$

把 g_2 除以 f_3(都看作 x_2 的多项式),得

$$u_3 g_2 = -(-2u_1 + 2x_1)f_3 + R_3,$$

其中

$$R_3 = (-2u_1u_2 + 2u_2x_1 + 2u_3^2)x_3 + u_1^2u_3 - u_3x_1^2 - u_3^3。$$

再把 R_3 除以 f_2(都看作 x_3 的多项式),得

$$(x_1 - u_1 - u_2)R_3 = (-2u_1u_2 + 2u_2x_1 + 2u_3^2)f_2 + R_2,$$

其中

$$\begin{aligned} R_2 &= -u_1u_3(-2u_1u_2 + 2u_2x_1 + 2u_3^2) \\ &\quad + (x_1 - u_1 - u_2)(u_1^2u_3 - u_3x_1^2 - u_3^3) \\ &= -u_3x_1^3 + (u_1u_3 + u_2u_3)x_1^2 \\ &\quad + (-2u_1u_2u_3 + u_1^2u_3 - u_3^3)x_1 \\ &\quad + (-u_1^3u_3 + u_1^2u_2u_3 - u_1u_3^3 + u_2u_3^3)。 \end{aligned}$$

最后把 R_2 除以 f_1(都看作 x_1 的多项式),得

$$R_2 = (u_3x_1^2 - 2u_1u_3x_1 + u_1^2u_3 + u_3^3)f_1 + R,$$

其中 R 恒等于 0。

于是在 $u_3 \neq 0$ 及 $x_1 - u_1 - u_2 \neq 0$ 的附加条件下(5)式得证。$u_3 \neq 0$ 及 $x_1 - u_1 - u_2 \neq 0$ 的意义可见第三节靠近末尾的部分。

例 3 求证:菱形对角线互相垂直。

选取坐标系,使菱形的一个顶点 A 为原点,一条边所在的直线为横轴,如图 18。

设 A、B 的坐标分别为 $A(0,0)$, $B(u_1,0)$。因 AB 与 BC 的长度应相等,故 C 点的坐标只能有一个为独立坐标,设 C 点

的坐标为(u_2, x_1)。D 点的两个坐标都不独立，且由 $AB /\!/ DC$，D 点的纵标与C 点的纵标相同，设 D 点的坐标为(x_2, x_1)。

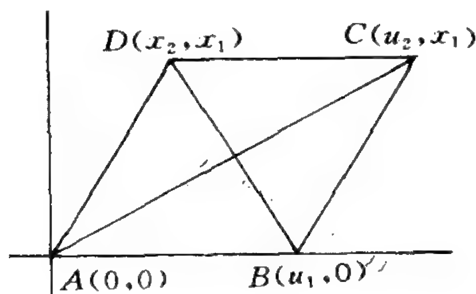


图 18

由 $AD /\!/ BC$，得

$$x_1(u_2 - u_1) - x_2 \cdot x_1 = 0。\tag{1}$$

由 $AB = AD$，得

$$u_1^2 - (x_2^2 + x_1^2) = 0。\tag{2}$$

(1)式中 $x_1 \neq 0$(否则 DC 与 AB 在同一直线上)，故可把(1)、(2)两式三角化如下：

$$f_1 = -x_2 + (u_2 - u_1) = 0,$$

$$f_2 = -(x_1^2 + x_2^2) + u_1^2 = 0。$$

终结 $AC \perp BD$ 为

$$g = x_1 \cdot x_1 + u_2 \cdot (x_2 - u_1) = 0。$$

把 g 除以 f_2，得

$$g = -f_2 + R_2,$$

其中 $$R_2 = -x_2^2 + u_2 x_2 - u_1 u_2 + u_1^2。$$

再把 R_2 除以 f_1，得

$$R_2 = (x_2 - u_1) f_1 + R,$$

其中 R 恒等于零。于是命题得证。

例 4 求证：等腰梯形底角相等。

如图 19，选取坐标系并设 A、B、C、D 各点的坐标。

由 $AD=BC$，得

$$f_1=x_1^2+u_3^2-(u_2-u_1)^2-u_3^2$$
$$=(x_1+u_2-u_1)(x_1-u_2+u_1)=0。$$

但因 AD 与 BC 不平行(梯形的定义)，故

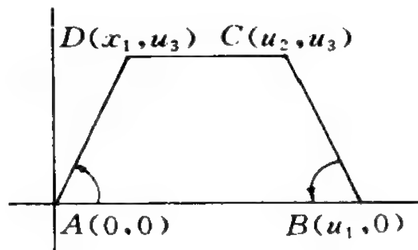


图 19

$$u_3\cdot(u_2-u_1)-x_1\cdot u_3\neq 0。$$

因此 $x_1+u_1-u_2\neq 0$，而上面的 $f_1=0$ 成为

$$f_2=x_1+u_2-u_1=0。$$

由第四节中的“7. 两角相等”，终结$\angle BAD=\angle CBA$ 可以表示为

$$g=u_3u_1[-u_1(u_2-u_1)]-u_3u_1(x_1u_1)$$
$$=-u_1^2u_3(x_1+u_2-u_1)=0。$$

因 $u_1\neq 0$(B 与 A 不重合)，$u_3\neq 0$(DC 与 AB 不在同一条直线上)，故由 $f_2=0$ 可以推得 $g=0$，而命题得证。

例 5 求证：三角形两条中线的交点，分顶点与对边中点成 2∶1。

选取两条中线的顶点所在的直线为横轴，其中一个顶点为原点，并设顶点 A、B、C 的坐标如图 20。由第四节中的“2. 一点是两点的中点”可得 BC 边的中点 D 和 AC 边的中点 E 的坐标。

设 AD 和 BE 的交点是 G。则由 A、G、D 共线和 B、G、E 共线可得

$$x_2(u_1+u_2)-x_1u_3=0,$$
$$x_2(u_2-2u_1)-(x_1-2u_1)u_3=0。$$

三角化得

$$f_1=u_1u_3[3x_1-2(u_1+u_2)]=0,$$

$$f_2 = x_2 \cdot (u_1 + u_2) - x_1 u_3 = 0。$$

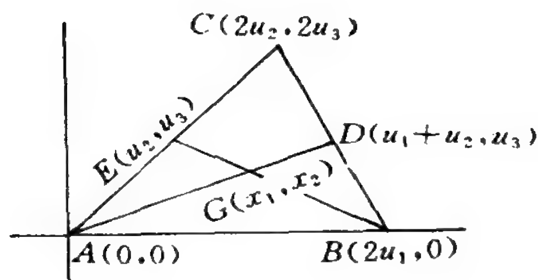


图 20

由第四节中的“10. 一点分两点成定比”，终结 G 分 A、D 成 2:1 和 G 分 B、E 成 2:1 可以表示为

$$\begin{cases} g_1 = x_1 - 2(u_1 + u_2 - x_1) = 0, \\ g_2 = x_2 - 2(u_3 - x_2) = 0; \end{cases}$$

$$\begin{cases} g_3 = x_1 - 2u_1 - 2(u_2 - x_1) = 0, \\ g_4 = x_2 - 2(u_3 - x_2) = 0。 \end{cases}$$

因 $u_1 \neq 0$（B 与 A 不重合），$u_3 \neq 0$（C 点不在 AB 直线上），故由 $f_1 = 0$ 可以推得 $g_1 = 0$, $g_3 = 0$。

$g_2 = 0$ 和 $g_4 = 0$ 都是 $g = 3x_2 - 2u_3 = 0$。

把 g 除以 f_2，得

$$(u_1 + u_2) g = 3f_2 + R_2,$$

其中

$$R_2 = u_3 [3x_1 - 2(u_1 + u_2)]。$$

再把 R_2 除以 f_1，得

$$u_1 R_2 = f_1 + R,$$

其中 R 恒等于 0。

因 $u_1 \neq 0$，于是在 $u_1 + u_2 \neq 0$ 的条件下，命题得证。

若 $u_1 + u_2 = 0$，则三角化以前的题设成为

$$f_1=-u_3x_1=0,$$
$$f_2=-3u_1x_2-u_3x_1+2u_1u_3=0。$$

而终结成为

$$\begin{cases}g_1=3x_1=0,\\ g_2=3x_2-2u_3=0;\end{cases}$$
$$\begin{cases}g_3=3x_1=0,\\ g_4=3x_2-2u_3=0。\end{cases}$$

又因 $u_1u_3\neq0$，故由题设 $f_1=0, f_2=0$ 也可推出终结 $g_1=0, g_2=0, g_3=0, g_4=0$。

例 6 求证：直角三角形斜边上的高是斜边上两线段的比例中项。

选取坐标系及 A、B、C、D 各点的坐标如图 21，其中 AD 是直角三角形 ABC 斜边 BC 上的高。

由 $AD\perp BC$，可得

$$x_2u_2+x_1(-u_1)=0。$$

由 B、D、C 三点共线，可得

$$x_2\cdot(-u_1)-(x_1-u_1)\cdot u_2=0。$$

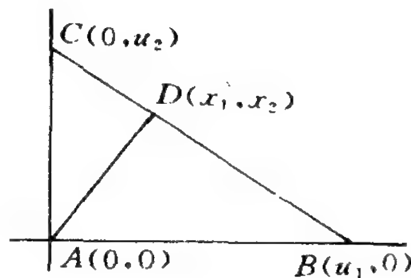


图 21

三角化后，得

$$f_1=u_1u_2^2-(u_1^2+u_2^2)x_1=0,$$
$$f_2=x_2u_2-x_1u_1=0。$$

终结 $AD^2=BD\cdot DC$ 可以表示为

$$g=[(x_1-u_1)^2+x_2^2]\cdot[x_1^2+(x_2-u_2)^2]-(x_1^2+x_2^2)^2=0。$$

把 g 除以 f_2(都看作 x_2 的多项式)，得

$$u_2^2g=[-2u_2^2x_2^2+(u_2^3-4u_1u_2x_1+u_1^2u_2)x_2+(5u_1u_2^2x_1-4u_1^2x_1^2+u_1^3x_1-2u_2^2x_1^2-2u_1^2u_2^2)]f_2+R_2,$$

其中 $\quad R_2=-4u_1(u_1^2+u_2^2)x_1^3+(u_1^4+6u_1^2u_2^2+u_2^4)x_1^2$

$$-2u_1u_2^2(u_1^2+u_2^2)x_1+u_1^2u_2^4。$$

再把 R_2 除以 f_1，得

$$R_2=[4u_1x_1^2-(u_1^2+u_2^2)x_1+u_1u_2^2]f_1+R,$$

其中 R 恒等于 0。

于是在 $u_2\neq0$（C 与 A 不重合）的附加条件下，命题得证。

例 7 求证：从三角形外接圆上一点到三角形三边所作垂线的垂足共线（这线称为西摩松线：Simson line）.

取$\triangle ABC$ 的 A 点为原点，AB 边所在的直线为横轴。设它的外接圆的圆心为 O，半径为 r。设 P 为外接圆上一点，它到$\triangle ABC$ 三边 AB、BC、CA 所作垂线的垂足分别为L、M、N。并设各点的坐标如图 22。

由 B、C、P 在圆上，得

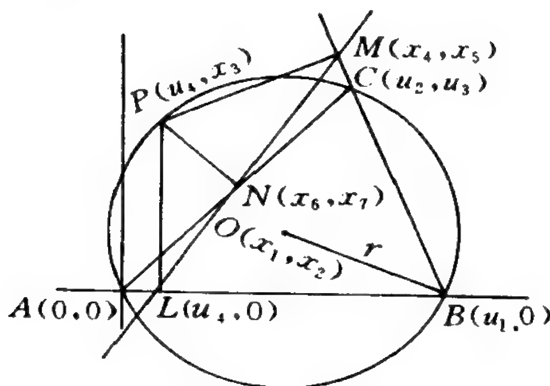


图 22

$$(u_1-x_1)^2+x_2^2=x_1^2+x_2^2,$$
$$(u_2-x_1)^2+(u_3-x_2)^2=x_1^2+x_2^2,$$
$$(u_4-x_1)^2+(x_3-x_2)^2=x_1^2+x_2^2。$$

由 $PM\perp BC$，$PN\perp AC$，得

$$(x_3-x_5)u_3+(u_4-x_4)(u_2-u_1)=0,$$
$$(x_3-x_7)u_3+(u_4-x_6)u_2=0。$$

由 M 在BC 上，N 在CA 上，得

$$x_5(u_2-u_1)-u_3(x_4-u_1)=0,$$
$$x_7u_2-x_6u_3=0。$$

经过三角化，得

$$f_1=2u_1x_1-u_1^2=0,$$
$$f_2=2u_3x_2+2u_2x_1-u_2^2-u_3^2=0,$$
$$f_3=x_3^2-2x_2x_3-2u_4x_1+u_4^2=0,$$
$$f_4=u_3^2(x_4-u_1)-x_3u_3(u_2-u_1)+(x_4-u_4)(u_2-u_1)^2=0,$$
$$f_5=x_5(u_2-u_1)-u_3(x_4-u_1)=0,$$
$$f_6=-x_6(u_2^2+u_3^2)+x_3u_2u_3+u_2^2u_4=0,$$
$$f_7=x_7u_2-x_6u_3=0。$$

终结 L、M、N 共线可以表示为

$$g=x_7(x_4-u_4)-(x_6-u_4)x_5=0。$$

把 g 除以 f_7(都看作 x_7 的多项式)，余式为(在某些条件下，下同)

$$R_7=x_6u_3(x_4-u_4)-u_2x_5(x_6-u_4)。$$

把 R_7 除以 f_6(都看作 x_6 的多项式)，余式为

$$R_6=u_2x_3x_5-u_3u_4x_5-u_3x_3x_4-u_2u_4x_4+u_3u_4x_3+u_2u_4^2。$$

把 R_6 除以 f_5(都看作 x_5 的多项式)，余式为

$$R_5=u_1u_2u_4x_4-u_3^2u_4x_4-u_2^2u_4x_4+u_1u_3x_3x_4-u_1u_2u_3x_3-u_1u_3u_4x_3+u_2u_3u_4x_3+u_2^2u_4^2-u_1u_2u_4^2+u_1u_3^2u_4。$$

把 R_5 除以 f_4(都看作 x_4 的多项式)，余式为

$$R_4=(u_2u_3-u_1u_3)x_3^2+(2u_1u_2^2-u_1^2u_2-u_2^3+u_1u_3^2-u_2u_3^2)x_3+u_2u_3u_4^2-u_1u_3u_4^2-u_1u_2u_3u_4+u_1^2u_3u_4。$$

把 R_4 除以 f_3(都看作 x_3 的多项式),余式为

$$R_3=(2u_2u_3x_3-2u_1u_3x_3)x_2+(2u_2u_3u_4-2u_1u_3u_4)x_1 +2u_1u_2^2x_3-u_1^2u_2x_3-u_2^3x_3+u_1u_3^2x_3 -u_2u_3^2x_3-u_1u_2u_3u_4+u_1^2u_3u_4。$$

把 R_3 除以 f_2(都看作 x_2 的多项式),余式为

$$R_2=(2u_1u_2x_3-2u_2^2x_3+2u_2u_3u_4-2u_1u_3u_4)x_1 +u_1u_2^2x_3-u_1^2u_2x_3-u_1u_2u_3u_4+u_1^2u_3u_4。$$

再把 R_2 除以 f_1,余式为 0。

考察过程中所需的条件,都能成立。因此命题得证。

例 7 因比较复杂,是借助于电子计算机而得到解决的。

通过上面的一些例子,可以看到机器证法的大致情况。这些例子,有的十分简单,用人工就可解决,有的略为复杂,由机器帮助解决。这些例子的机器证法,虽然看起来反而比传统证法更为复杂,但是如前所述,机器证法是按一定的步骤进行的,不象传统证法没有一定的规律可循,对于每一个不同的问题都要另费一番思考。

机器证法有一定的步骤,不怕复杂,所以可以用机器证法解决许多难题。这些难题,用传统的综合方法来证,因为不易找到思路,所以十分困难。

例如,有一个著名难题,叫做 Morley 定理,是说任意三角形 ABC 中,一个角的一条三等分线,与和它相邻的角的三等分线相交,象这样所得的交点 P、Q、R 是一个正三角形的三个顶点(如图 23)。

用传统方法证明 Morley 定理,非常艰难。用机器方法证明 Morley 定理,跟证明其他定理一样,有一定的步骤,只要按步进行就可以了。

进一步的研究表明,任意三角形中,一个角的三等分线,与和它相邻的角的三等分线相交,按一定的规则选取交点,共可组

成 27 个三角形，在这 27 个三角形中，一定有 18 个三角形是正三角形。*

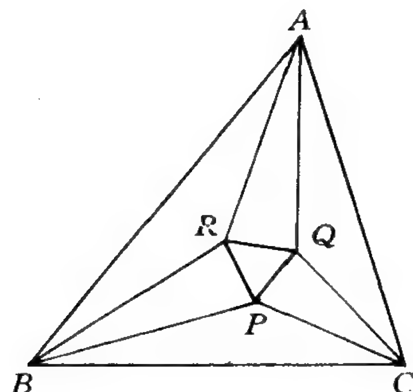


图 23

用机器方法同样容易证明这个更一般的 Morley 定理。在证明过程中，不止一次地出现了关于 12 个变量的含有一千多项(有的有 1960 项)的多项式。象这样的问题，不用机器而用人工处理，当然是非常困难的。

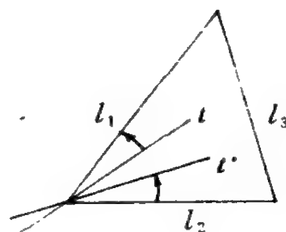
---

* 27 个三角形的组成如下。

设三角形的三边为 l_1, l_2, l_3。用$\angle(l_1, l_2)$表示 l_1, l_2 所成的角(要考虑旋转方向)，我们规定能使 $3\angle(t, l_1)$ 等于 $\angle(l_2, l_1)$，或等于 $\angle(l_2, l_1)+180°$，或等于 $\angle(l_2, l_1)+360°$ 的 t 为这个角的主三等分线。这样的主三等分线有三条。

对于每一条主三等分线 t，有一条相应的能使 $\angle(l_2, t')=\angle(t, l_1)$的副三等分线 t'。

$\angle(l_1, l_2)$的任一条主三等分线，与$\angle(l_2, l_3)$的任一条主三等分线的相应副三等分线相交，得一交点，$\angle(l_2, l_3)$的这条主三等分线与$\angle(l_3, l_1)$的任一条主三等分线的相应副三等分线相交，得一交点，$\angle(l_3, l_1)$的这条主三等分线与$\angle(l_1, l_2)$的所取的主三等分线相应的副三等分线相交，得一交点。三个交点组成一个三角形。这样共组成 27 个三角形。



六、回到前面的古老问题

说明了机器证法的大致情况并举了一些例子以后，我们再回到第一节所说的古老问题：

求证：两条内分角线相等的三角形是等腰三角形。看看怎样用机器证法来证明这个命题。

但是，对于这个古老问题来说，上面所说的方法还不够用，需要另外想一些办法。

这是因为，上面所说的方法，适用于能把命题的题设部分和终结部分都化为等式的情况，而不能用于化为不等式的情况。

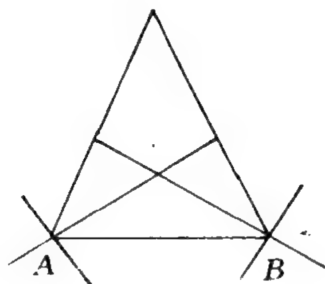


图 24

在这个古老问题中，按照第四节“8．点在分角线上”所列的等式，只能确定一点是在$\triangle CAB$的$\angle CAB$的分角线上，而不能确定究竟是在内分角线上还是在外分角线上。但这个古老问题是要证明两条内分角线相等的三角形是等腰三角形，而对于两条外分角线相等的三角形来说，情况就大不一样，（见下节）。因此，在这个古老问题中，一点究竟在内分角线上，还是在外分角线上，有着决定性的重要关系。但是这一点却不能单独用第四节“8．点在分角线上”所列的等式来表示，而需要另外添加一个不等式。

幸好，在这个古老问题中，所要添加的不等式是在题设部分，而不是在终结部分。这样，上面所说的机器证法只要稍加改进，就仍然能够适用。如果一个命题，在它的终结部分，需要用不等式来表示，那么上面所说的机器证法就无能为力了。

这就是说，上面所说的机器证法，**不能适用于终结部分需要用不等式来表示的命题**（也就是涉及希尔伯脱 Hilbert 公理体系中有关次序公理的问题）。例如，上面所说的机器证法，不能用来证明象下面这类的命题："三角形两边之和大于第三边"；"在一个三角形中，大角所对的边也大"；等等。

现在我们来看怎样改进一下上面所说的机器证明方法，可以用来证明这个古老问题。

如图 25，取 AB 的中点为原点，AB 所在直线为横轴，且以 AB 的长度的一半为单位长度，则 A, B 的坐标为 $A(-1,0)$，B（1，0）。

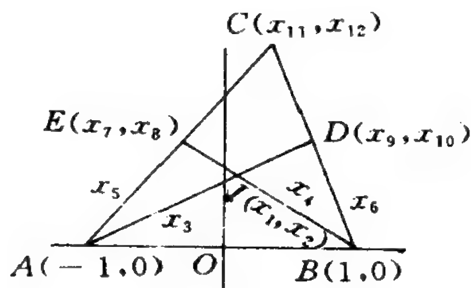


图 25

设分角线 AD 和 BE 相交于 I，并设 I 的坐标为 $I(x_1, x_2)$。设两条分角线 AD、BE 的斜率分别为 x_3、x_4，$\triangle ABC$ 的两边 $ACBC$ 的斜率分别为 x_5、x_6，E 点的坐标为 E（x_7、x_8），D 点的坐标为 D（x_9、x_{10}），C 点的坐标为 C（x_{11}，x_{12}）。

由 AI 的斜率为 x_3，可得

$$x_3 + x_1 x_3 - x_2 = 0。\qquad (1)$$

由 BI 的斜率为 x_4，可得

$$x_1 x_4 - x_4 - x_2 = 0。\qquad (2)$$

由 AB 旋转到 AI 所成的角等于 AI 旋转到 AC 所成的角，可得

$$x_5 - x_3^2 x_5 - 2x_3 = 0。\qquad (3)$$

由 BA 旋转到 BI 所成的角等于 BI 旋转到 BC 所成的角，可得

$$x_6 - x_4^2 x_6 - 2x_4 = 0。\qquad (4)$$

由 AE 和 AC 的斜率都等于 x_5，可得

$$x_8 - x_5 x_7 - x_5 = 0, \tag{5}$$

$$x_{12} - x_5 x_{11} - x_5 = 0。 \tag{6}$$

由 BD 和 BC 的斜率都等于 x_6，可得

$$x_{10} - x_6 x_9 + x_6 = 0, \tag{7}$$

$$x_{12} - x_6 x_{11} + x_6 = 0。 \tag{8}$$

由 A、I、D 共线，可得

$$x_{10} + x_1 x_{10} - x_2 x_9 - x_2 = 0。 \tag{9}$$

由 B、I、E 共线，可得

$$x_8 - x_1 x_8 + x_2 x_7 - x_2 = 0。 \tag{10}$$

此外，又由分角线 AD 与 BE 相等，可得

$$x_{10}^2 + x_9^2 + 2x_9 - x_8^2 - x_7^2 + 2x_7 = 0。 \tag{11}$$

上面所列的等式（3）、（4），只能表示 D 在$\angle CAB$ 的分角线上，E 在$\angle CBA$ 的分角线上，而不能表示 AD、BE 究竟是内分角线还是外分角线。

为了表示 AD、BE 都是内分角线，我们添设两个条件：

1. AD 和 BE 的斜率的符号相反，这说明 AD 和 BE 同是内分角线或同是外分角线。

2. AD 和 BE 的交点 I 与 C 点在 AB 的同侧。

这两个条件合起来就可确定分角线 AD 和 BE 同是内分角线。

条件 1 用式子表示就是

$$x_3 x_4 < 0。$$

为了用等式而不是用不等式来表示这个条件，我们可以设一个实数 u，而把上面的不等式写成下面的等式：

$$x_3 x_4 = -u^2。 \tag{12}$$

条件 2 可以表示为

$$x_2 x_{12} > 0。 \qquad (13)$$

本题的终结部分原来应是 $AC = BC$。但是，为了简便，可以用"I 在纵轴上"来代替。这是因为，I 在 AB 的中垂线上，则 $IA = IB$，$\angle IAB = \angle IBA$，即可得$\angle CAB = \angle CBA$，从而 $AC = BC$。因此，终结部分可以表示为

$$x_1 = 0。 \qquad (14)$$

我们先把条件（13）暂且不管，留作以后讨论时用，而把（1）～（12）式三角化如下（利用机器）：

$f_1 = -u^4 x_1 \ (1 - x_1^2)^4 \cdot f_0 = 0,$

其中 $f_0 = \ (u^2 - 1)^2 \ (u^2 - 2) \ (1 - x_1^2) \ - 4;$

$f_2 = x_2^2 - u^2 + u^2 x_1^2 = 0;$

$f_3 = x_3 + x_1 x_3 - x_2 = 0;$

$f_4 = x_1 x_4 - x_4 - x_2 = 0;$

$f_5 = (1 + x_1)[(1 + x_1 - u^2 + u^2 x_1) \cdot x_5 - 2x_2] = 0;$

$f_6 = {x_1}^2 x_6 - 2x_1 x_6 + x_6 - u^2 x_6 + u^2 x_1^2 x_6 - 2x_1 x_2 + 2x_2 = 0;$

$$f_7 = 3x_2 x_7 - x_1^2 x_2 x_7 + x_2 - 3x_1^2 x_2 + 2x_1 x_2 x_7 - u^2 x_2 x_7 + u^2 x_1^2 x_2 x_7 - 2x_1 x_2 + u^2 x_2 - u^2 x_1^2 x_2 = 0;$$

$f_8 = x_8 - x_7 x_5 - x_5 = 0;$

$$f_9 = x_1^2 x_2 x_9 - 3x_2 x_9 - 3x_1^2 x_2 + x_2 + 2x_1 x_2 x_9 + u^2 x_2 x_9 - u^2 x_1^2 x_2 x_9 + 2 \cdot x_1 x_2 + u^2 x_2 - u^2 x_1^2 x_2 = 0;$$

$f_{10} = x_{10} - x_9 x_6 + x_6 = 0;$

$f_{11} = x_2(1 - x_1^2)[x_{11}(1 - u^2) - x_1(1 + u^2)] = 0;$

$f_{12} = x_{12} - x_5(x_{11} + 1) = 0。$

下面我们实际上不必继续进行，因为在 f_1 中已经含有因式 x_1，而 f_1 等于零。如果我们能够利用这些题设条件，加上题设条件（13），证明 f_1 中的其他因式都不等于零，那么就能得出终结 $x_1 = 0$。

在 f_1 中，有因式 $-u^4, x_1, (1-x_1^2)^4$，以及 $f_0=(u^2-1)^2(u^2-2)(1-x_1^2)-4$。

我们来考察除了 x_1 以外的其他因式能否为零，如果除了 x_1 以外的其他因式都不能为零，那么由 $f_1=0$ 就可推出我们希望得到的终结 $x_1=0$。

由（12），$x_3x_4=-u^2$，而 x_3，x_4 分别是分角线 AD 和 BE 的斜率（如图 25），所以 $x_3x_4\neq 0$，因而 $u\neq 0$。

又在图 25 中，x_1 表示两条分角线 AD 和 BE 的交点 I 的横坐标，所以 $x_1\neq -1$，$x_1\neq +1$，否则 A 角或 B 角就将成为 180° 而不能构成 $\triangle ABC$。从而可得 $1-x_1^2\neq 0$。

由 $f_2=0$，可得 $1-x_1^2=\dfrac{x_2^2}{u^2}$，因此，$1-x_1^2>0$。

由 $f_5=0$，因 $1+x_1\neq 0$，可得 $x_5=\dfrac{2x_2}{1+x_1-u^2+u^2x_1}$；又由 $f_{11}=0$，因 $x_2\neq 0$（x_2 是分角线 AD 的 BE 的交点 I 的纵坐标），$1-x_1^2\neq 0$，可得 $x_{11}=\dfrac{x_1\ (1+u^2)}{1-u^2}$；代入 $f_{12}=0$，即得 $x_{12}=\dfrac{2x_2}{1-u^2}$，从而可得 $x_2x_{12}=\dfrac{2x_2^2}{1-u^2}$。

但由条件（13），$x_2x_{12}>0$，于是可得 $1-u^2>0$，即 $u^2<1$。

综上所述，在条件（12）、（13）成立的情况下，即在 AD、BE 都是内分角线的情况下，f_1 中的除了 x_1 以外的其他因式 $-u^4\neq 0$，$(1-x_1^2)^4\neq 0$，且 $f_0=\ (u^2-1)^2\ (u^2-2)(1-x_1^2)\ -4<0\ ((u^2-1)^2>0,\ (1-x_1^2)\ >0,\ u^2-2<1-2<0)$，即恒为负值而 $f_0\neq 0$，因此由 $f_1=0$ 可以得到 $x_1=0$。

这样就证明了我们所要证明的古老问题：

两条内分角线相等的三角形是等腰三角形。

七、“两条外分角线相等的三角形是等腰三角形”，这个猜想成立吗？

上节证明了古老问题：“两条内分角线相等的三角形是等腰三角形。”

由此很容易引起我们产生下面的猜想：“两条外分角线相等的三角形是等腰三角形”。这个猜想能不能成立呢？

举一个反例*，就可证明这个猜想不能成立！

设在图 26 的$\triangle ABC$中，$\angle A=132^\circ$，$\angle B=12^\circ$，$\angle C=36^\circ$。外角 CAF 的分角线 AD 与对边 BC 交于 D，外角 ABG 的分角线 BE 与对边 AC 交于 E，那么

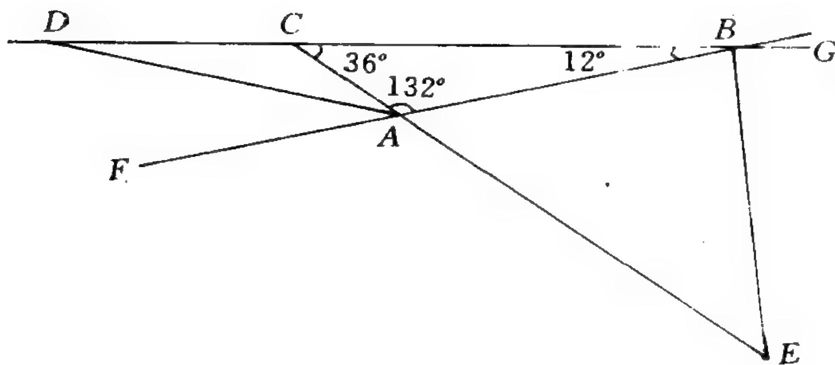


图 26

$$\angle CAD=\frac{1}{2}(180^\circ-132^\circ)=24^\circ,$$

$$\angle ACD=180^\circ-36^\circ=144^\circ,$$

$$\therefore\quad \angle D=12^\circ。$$

---

* 见《数学通报》，1983 年 1 月号问题及解答。

又 $\angle ABE=\frac{1}{2}(180^\circ-12^\circ)=84^\circ$，

$\angle BAE=180^\circ-132^\circ=48^\circ$，

∴ $\angle E=48^\circ$。

因此 $AD=AB$， $AB=BE$。

就是说，$\triangle ABC$ 的两条外分角线 AD 和 BE 相等，而 $\triangle ABC$ 并不是等腰三角形。

再从更一般的情况来作直观的考察。

设$\triangle ABC$ 的 AB 边固定，BC 边所在的射线也固定（$\angle ABC$ 是小于 60°的一个定角 α），如图 27。但 C 点的位置并不固定，可以在这条射线上移动，我们研究 C 点在这条射线上移动时所产生的情况。BC 所在的射线既已固定，那么外角 ABG 的分角线 BE（直线）也就确定了。

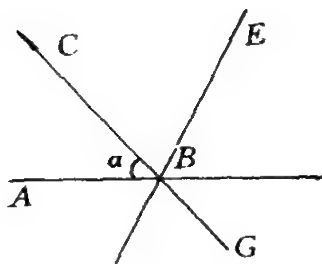


图 27

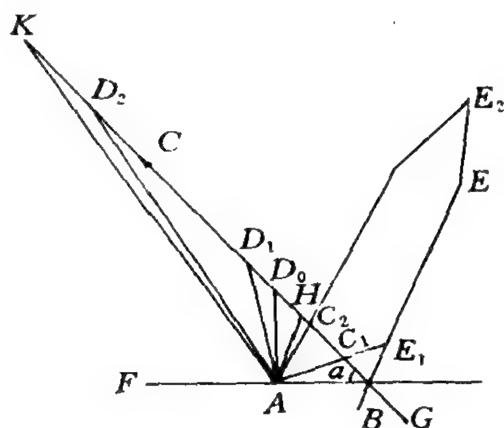


图 28

如果 C 点接近 B 点，如图 28 中的 C_1，那么外分角线 BE_1 接近于零，而外分角线 AD_1，接近于 AD_0，D_0 是从 A 所作 AB 的垂线与 BC 的交点。因此，在这个情况下，$AD_1>BE_1$。设过 A 所作 BE 的平行线与 BC 交于 H，那么当 C 点接近于 H 点时，

如图中的 C_2，外分角线 BE_2 无限增大（图中有歪曲），而外分角线 AD_2 接近于 AK，K 是外角 HAF 的分角线与 BC 的交点。因 $\angle KAF=\frac{1}{2}\angle HAF=\frac{1}{2}\left[\alpha+\frac{1}{2}(180^\circ-\alpha)\right]=45^\circ+\frac{\alpha}{4}\neq\alpha$（因 $\alpha\neq60^\circ$），所以 AK 不平行于 BC，因而 AD_2 是有限值。因此，在这个情况下，$AD_2<BE_2$。

由此可见，C 点从 B 到 H 连续运动时，由 $AD_1>BE_1$ 连续变化到 $AD_2<BE_2$，中间应至少有一处能使 $AD=BE$。

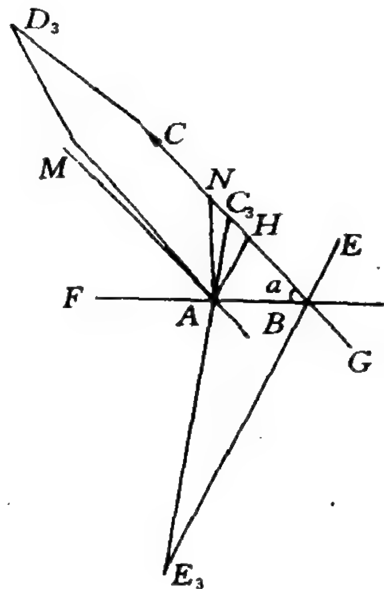


图 29

C 点超过 H 而接近于 H 时，如前所说，仍有 $AD_2<BE_2$。如图 29，作 $AM /\!/ BC$，且作 AF 对于 AM 的对称线 AN，交 BC 于 N。因 $\angle FAN=2\alpha$，$\angle FAH=180^\circ-\alpha$，且 $\alpha<60^\circ$，故 $\angle FAN<\angle FAH$。就是说，N 点与 B 点在 H 点的异侧。当 C 点接近 N 点时，如图中的 C_3，AC_3 就不与 AH 接近，外分角线 BE_3 是有限值。而 $\angle C_3AF$ 的分角线接近于 AM，因而外分角线 AD_3 无限增大（图中有歪曲）。因此，在这个情况下，$AD_3>BE_3$。

由此可见，C 点从 H 到 N 连续运动时，由 $AD_2<BE_2$ 连续变化到 $AD_3>BE_3$，中间应至少有一处能使 $AD=BE$。

综上所述，在 AB 边固定且 $\angle ABC$ 固定（小于 60°）的前提下，至少有两个 C 点（一在 B、H 之间，一在 H、N 之间）能使 $\triangle ABC$ 的外分角线 AD 与 BE 相等，而不是只有唯一的一个 C 点（AB 的中垂线与 BC 的交点）。由此得出结论，两条外分角线相等的三角形不一定是等腰三角形。这就说明我们原来的猜想

“两条外分角线相等的三角形是等腰三角形”不能成立。

从我们的机器证明来看，上节关于两条内分角线所列的(1)～(12)式以及由此推出的 $f_1=0, f_2=0, \cdots, f_{12}=0$ 等式都能适用于本节关于两条外分角线的情况，所不同的只是要把原来的(13)式 $x_2 \cdot x_{12}>0$ 改为

$$x_2 \cdot x_{12}<0。\qquad (13')$$

于是，由 $x_2 x_{12}=\frac{2x_2^2}{1-u^2}$，可得 $\frac{2x_2^2}{1-u^2}<0$，即 $u^2>1$。因而 f_1 中的因式 $f_0=(u^2-1)^2(u^2-2)(1-x_1^2)-4$（其中 $1-x_1^2=\frac{x_2^2}{u^2}>0$）就不一定恒为负值，而可能为 0。这时，由 $f_1=0$ 就不能推出 x_1 一定为 0，就是说，$\triangle ABC$ 不一定是等腰三角形。

由于 $1-x_1^2=\frac{x_2^2}{u^2}$，即 $u^2=\frac{x_2^2}{1-x_1^2}$，所以

$$f_0=(u^2-1)^2(u^2-2)(1-x_1^2)-4=0$$

可以改写为

$$x_2^6-4x_2^4(1-x_1^2)+5x_2^2(1-x_1^2)^2$$
$$-2(1-x_1^2)^3-4(1-x_1^2)^2=0。\quad (x_1^2<1)$$

我们知道，(x_1, x_2) 是分角线 AD 与 BE 的交点 I 的坐标。当 I 点的坐标满足上述方程时，I 点在某一个轨迹上。因此，从 $f_1=-u^4x_1(1-x_1^2)^4f_0=0$，可以得出：或者 $x_1=0$，即 $\triangle ABC$ 为等腰三角形；或者 $f_0=0$，即分角线 AD 与 BE 的交点 $I(x_1, x_2)$ 在上述方程所表示的轨迹上。

这就说明了，以 AB 为一边，可以有无数个 $\triangle ABC$，其中 $AC \neq BC$，而外分角线 AD 等于外分角线 BE。这些三角形中，外分角线 AD 与 BE 的交点 I 的轨迹是一个近似于椭圆的图形（如图 30），与过 A 而垂直于 AB 的直线相切，又与过 B 而垂直于 AB 的直线相切，它的最高点与 A、B 两点恰好组成一个正三

角形。按通常理解的三角形来说，这个轨迹要除去 A、B 两点以及 AB 的中垂线与轨迹相交的两点。

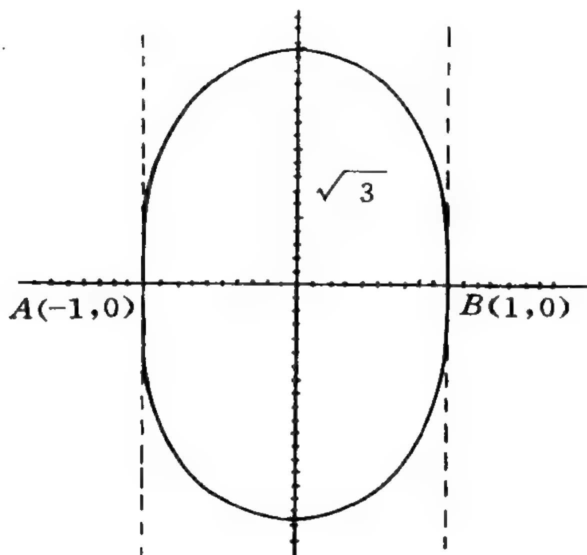


图 30

本书封面上用实线画出的一个三角形，不是等腰三角形，但它的两条外分角线（也用实线画出）相等。这两条外分角线所在直线的交点就在这样的轨迹上。

最后，再说一下一条内分角线与一条外分角线相等的三角形。研究结果表明，这样的三角形也不一定是等腰三角形。与两条外分角线相等的三角形的情况相仿，也有下面的情况：以 AB 为一边，可以有无数个 $\triangle ABC$，其中 $AC \neq BC$，而内分角线 AD 等于外分角线 BE（或外分角线 AD 等于内分角线 BE）。这些三角形中，AD 与 BE 的交点 $I(x_1, x_2)$ 的轨迹是一个近似于椭圆的图形，如图 31 中的右图（或左图），它的方程是

$$x_2^6 + 4x_2^4(x_1^2 - 1) + 5x_2^2(x_1^2 - 1)^2 + 2(x_1^2 - 1)^3 - 4(x_1^2 - 1)^2 = 0。\quad (x > 1 \text{ 或 } x < -1)$$

这个图形，与过 B 及过 P 而垂直于 AB 的直线相切于 B 及 P（或与过 A 及过 Q 而垂直于 AB 的直线相切于 A 及 Q），其中 P（或

Q)是在 AB(或 BA)的延长线上且与 AB 的中点相距$\frac{\sqrt{3}}{2}AB$ 的点。按通常理解的三角形来说,轨迹上也应除去一些特殊的点。

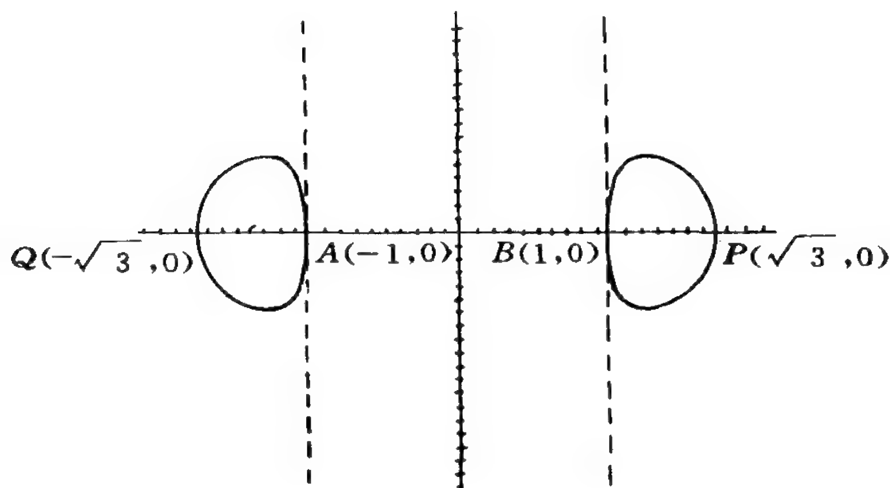


图 31

*　　　*　　　*

通过上面所说,不难看到,机器证明有一定的优越性。本文只介绍一些大意和具体问题,不能过多谈到理论性问题。当然,这个问题还处在开创阶段,有待于进一步的发展与完善。

解方程器或 SOLVER 软件系统概述*

一、引　　言

方程求解，无疑是数学通向实际应用为经济建设服务的主要途径之一。不仅如此，方程求解又是推动数学发展的巨大动力，举例来说，线性联立方程组的求解导致了正负数概念的引入以及矩阵、线性变换、线性空间等理论的建立。高次方程求解又导致了复数概念的引入以及能否用根式求解的 Galois 理论以至有关群论的创立。至于高次联立方程组的求解，更与当代蓬勃发展中的代数几何关系密切。事实上代数几何研究的主要对象代数簇，无非是高次联立方程组的全体解答所构成的图象而已.

方程求解是我国古代数学的传统中心内容之一，两千多年来的成就为世界数学的发展作出不可磨灭的贡献。在西汉早年即已出现的《九章算术》，其中第八章方程，即是线性联立方程组的解法，并包括了正负数的概念与移项运算等法则。依据魏晋时期刘徽的《九章注》，解法用各种消去法，当时在筹算板上的运算过程，实质上与现代用矩阵逐步进行计算的过程无异，这种消去法，直至 1828 年才重见于 Gauss 的著作，现在通称为 Gauss 消去法，参阅［WU5］中的文献［Gl］。至于高

---

* 本文摘自《数学的实践与认识》，1986 年第 2 期。

次方程解法，则自《九章》中即已出现的开平方、开立方、与开带从平方（即解二次方程），到唐初王孝通的开带从立方（即解三次方程），以迄宋代1247年秦九韶的《数学九章》，已可求任意高次方程的数值解，其方法基本上与1819年出现的Horner法相同。在宋元时期，又创立了天元术，相伴而生的是多项式概念，表达方法，运算法则以及一般消去法的建立，到元朱世杰的《四元玉鉴》，已可解多至四个未知数的高次联立方程组。只是由于我国古代不用笔算而用筹算，计算须在筹算板上进行，才只能局限于四个未知数。就其理论与方法的实质而论，应是可以推行于任意高次联立方程组的。元代以后，我国古代数学的发展卒然中断，经典文献也散失者多保存者少。但即使从幸存至今的文献来看，方程求解的发展过程，还是脉络分明，有书可稽的。

在数学中，线性联立方程组的求解是一种线性问题，而高次联立方程组的求解则是一种非线性的问题，两者难易之差何啻霄壤。前者在数学中已相当成熟，而后者则不论是理论（如代数几何）或是具体求解，国外似乎都还在初期的探索阶段。就作者初步的调查所知，国外解高次联立方程组的方法，或则失之于实际上难以实施，或则失之于理论上的缺陷，不能保证求得解答的全体，甚至在方法上也有根本可疑之处。与之相反，作者遵循我国古代数学传统的思想与方法，结合现代数学中的某些理论（主要是Ritt的理论，参阅文献［R1，2］），提出了Ritt原理并提炼成多项式组的零点分解定理（见文献［WU2，4］），据之以得出了彻底解决高次联立方程组求解的方法。我们的方法，在理论上可以确切判断一个高次联立方程组有无解答，有否无穷多解，以及在恰有有限多个解答时，求出这些解答的全部而无增无漏。在实际上，这一方法又是切实可行的，我们已在Dual System83/20高档微型机上，用for-

tran 77 编成程序，并用之以解决从实际与理论中出现的若干形形色色的问题。这些实例的具体计算过程，以及相应的程序与说明书，统称之为本文标题所示的解方程器或 SOLVER 软件系统，并已灌制在一片 Dysan 的 8″软磁盘中。使用者可据之以移植到其他类型的机器，或用他种语言（basic，pascal 等）译写，并以之解决实际与理论中出现的多方面的问题。

由于本软件系统的理论依据在作者关于机器证明的多篇论著中已有所阐述（见［WU1—4］），故本文对此将只作一些概括叙述，而着重于程序的编写与使用方法上。至于各种实例的说明与具体计算，则将在另一文（［WU5］）中详细叙述。

在作者使用机器以建立这一软件系统时，曾经得到许多同志的无私帮助。其中特别应该提到夏启圣同志，以及计算所八室胡增炎、文逸等同志，与计算中心的李平、唐友三同志等，作者谨向他们表以深切谢意.

二、方法概述

所有多项式的变量，将固定为 x_1，…，x_n，系数则可以是整数，实数或复数.

对于一个多项式 A，其中真正出现的 x_i 的最大下标 i 称为 A 的类数，记作 cls（A）. 在 A 为一个非 0 常数时，类数定为 0，即 cls（A）＝0。若 cls（A）＝$c>0$，则 A 中 x_c 的最高次数简称为 A 的次数，记作 deg（A）。记 A 的项数为 t，真正出现的变量个数为 v。将 A 依通常形式写出时，将其中的变量按其出现次序排成一个数组记为 var（A）时，其中 x_c 的位置设为 p，则 A 附有一个数组

$$\mathrm{ind}(A)=[t \quad v \quad c \quad d \quad p],$$

这里 v 是 var(A)的个数，$c=\mathrm{cls}(A)$，$d=\deg(A)$。一个多项

式 A，因而可用三个数组来表示：一个一维数组 var(A)，一个二维数组 exp(A)，其中第 i 行第 j 列的元素为 A 中第 i 项 $x_{\mathrm{var}(j)}$ 的次数，以及一个一维数组 cof(A)，其中第 i 个元素为 A 中第 i 项的系数。

多项式 A 的这一表达形式，将称为 A 的矩阵形式，显然这种形式与 A 以通常方式写出时的形式有关。

在 $\mathrm{cls}(A)=c>0$，$\deg(A)=d$ 时，A 中 x_c^d 的系数本身是一个多项式，称为 A 的初式，记作 init(A)。这一概念在整个方法中扮演着重要的角色。

考虑一个由有限多个非 0 多项式 A_i 所成的多项式组

$(PS)A_1, A_2, \cdots, A_r$。

由 $A_i=0$，$i=1, \cdots, r$ 所构成的方程组将简记为 $(PS)=0$。这一方程组的所有复数值解或 (PS) 的零点构成复 n 维线性空间 C^n 中的一个集合，记作 Zero(PS)。如果另有一非 0 多项式 G，则 Zero(PS) 中使 $G\neq 0$ 的部分将记作 Zero(PS/G)。

所谓方程求解，意指给定一个方程组 $A_i=0$，将其左边的多项式 A_i 构成一个多项式组 (PS) 时，求出 Zero(PS) 即须判明 Zero(PS) 是否是空集（原方程组无解），非空时是否是有限集，在无限集时求得这些解的参数及其依赖情况，以及有限集时给出这些解或零点的全体。

我们的解法实质上与《九章算术》方程术（刘徽《九章注》给出了三种不同的消去法）或近世所称 Gauss 消去法无异，即设法将原来的方程组化成下面的三角阵形式：

$$
\begin{aligned}
&E_1(x_1)=0,\\
&E_2(x_1, x_2)=0,\\
&\cdots\\
&E_n(x_1, x_2, \cdots, x_n)=0,
\end{aligned}
$$

于是可以逐个地依次解出 x_1，x_2，$\cdots$，x_n。自然，这里的消

去过程远比线性的情况要复杂，而且也可以出现几种不同的情形（这在线性时也是一样的），即

情形一　化得一个常数多项式 $E_0=0$。

情形二　化得若干个 $E_i(x_1,\cdots,x_i)=0$，但并非每一 E_i 都出现。不论何时，我们将改写 $E_i=0$ 为下面的形式：

$$C_1(x_1,\cdots,x_{p1})=0,$$
$$C_2(x_1,\cdots,x_{p_1},\cdots,x_{p_2})=0,$$
$$\cdots$$
$$C_s(x_1,\cdots,x_{p_1},\cdots,x_{p_2},\cdots,x_{p_s})=0,$$

其中 $\mathrm{cls}C_i=p_i, 0<p_1<p_2<\cdots<p_s$。又对任意 $j<i$，C_i 中 x_{p_j} 的最高次数都少于 $\deg C_j$，即 C_j 中 x_{p_j} 的最高次数，简称为（性质 P）。此外，记 C_i 的初式为 I_i，J 为诸 I_i 的乘积。又诸 C_i 所构成的多项式组为（CS）时，有

$$\mathrm{Zero}\,(PS)=\mathrm{Zero}(CS/J)+\mathrm{Sum}_i\mathrm{Zero}(PS_i), \qquad (Z)$$

其中(PS_i ）为将 I_i 添入（PS）所得扩大了的多项式组，+ 与 Sum 指集合的并。

在情形一中，记单个常数多项式 E_0 所成的多项式组为（CS），则也有

$$\mathrm{Zero}\,(PS)=\mathrm{Zero}\,(CS)=\text{空集},$$

可视为公式(Z）的一个特例，不论何时多项式组（CS）都称为（PS）的特征组。

从（Z）式出发继续进行，可获得更一般的公式：

$$\mathrm{Zero}\,(PS)=\mathrm{Sum}_k\,(AS_k/R_k), \qquad (Z)^*$$

其中每一（AS_k）都是一多项式组，有着类似于（CS）的那种形式，并满足性质 P，而 R_k 是一非 0 多项式。公式（Z）与（Z)* 中的那些多项式与多项式组都可从原来的（PS）依确定步骤得出。

上面这些结果我们称之为 Ritt 原理与零点分解定理，其证明将从略，读者可参阅文末所附 Ritt 与作者的有关文献。

对我们重要的是从多项式组（PS）求得特征组（CS）的具体过程，如下面的格式所示：

$$
\begin{array}{ccccccc}
(PS) = (PS_1) & \to & (PS_2) & \to \cdots \to & (PS_m) & & \\
\cup & & \cup & & \cup & & \\
(BS) = (BS_1) & & (BS_2) & & (BS_m) & & \text{(SCH)} \\
(RS_1) & & (RS_2) & & (RS_m) & &
\end{array}
$$

在这一格式中，每一（BS_i）都是（PS_i）的一个最大可能的子组，具有类似于（CS）的形式并满足性质（P）。每一（RS_i）为（PS_i）中除去（BS_i）后的多项式对（BS_i）中多项式依 x_i 的相反次序逐个相除所得余式凡非 0 者所成的多项式组，又每一（PS_i）为将前面所得 RS_{i-1} 添入（PS_{i-1}）后所得扩大了的多项式组，即

$$(PS_i) = (PS_j) + (RS_j), j = i - 1. \qquad \text{(PR)}$$

这一过程必然终止于某一 m，使相应的（RS_m）为空集。于是（BS_m）即为所求的(CS)。这些（BS_i）各称为（PS_i）的基组。

在具体计算的过程中，即使原来的（PS）只有很少几个多项式，每个多项式只有很少几项，也可能迅速膨胀，使多项式的项数增至几千甚至一二万以上。特别是在上面的格式(SCH）中，（PS_i）中多项式的个数都逐步增加，这种膨胀更容易发生，为了尽可能避免或至少延缓这种过份膨胀的发生，我们对上面引入的概念与过程略作修改，而不致影响最后的结果。

第一种修改是关于基本概念方面的。在原来关于（CS）的定义中，要求具有性质（P），现将（P）的要求减弱为

（P*）对任意 $j<i$，只须 xp_j 真正出现在 C_i 的初式中，

C_i 对 x_{p_j} 的最高次数即小于 $\deg C_j$。

具有性质（P*）的（CS）将称为弱型的，原来具有性质（P）的（CS）称为强型的。同样对于基组，特征组等的类似概念都将以弱型或强型来区别其所具有的相应性质。

第二种修改是关于格式（SCH）的。原来要求（PS_i）依（PR）一式来作出，这样（PS_i）中的多项式个数将越来越多，容易超出机器可以容纳的范围。为此可以把（PS_i）的作法改为以下二者之一：

$$(PS_i) = (PS) + (RS_j),\ j = i - 1, \qquad (\mathrm{PR}^*)$$

或

$$(PS_i) = (BS_j) + (RS_j),\ j = i - 1。 \qquad (\mathrm{BR})$$

依据这种修改形式后所得的(CS)，将称为原来（PS）的半特征组。

第三种修改是关于计算中出现的各个基组（BS_i）的。原来关于零点分解定理的严格理论，要求对诸（CS）中的多项式进行因子分解。因子分解本来是计算机科学中的一个困难问题，特别是在我们考虑的场合需要在某些代数扩域中来进行时更是如此，尽管理论上完全可以做到（参阅文 [WU4])，但将为此而增加许多计算量。事实上，对于方程求解来说，因子分解并不是必要的。但为了尽可能减少计算中出现的多项式，我们采取了下面的修改方式：对于在格式（SCH）中出现的任一基组（BS_i）中的任一多项式，如果其类为 c，即将这一多项式中各 x_c^e 的系数（本身也是多项式）的最大公约式约去。注意最大公约式的求法与因子分解不同，是容易进行的。

至此，可以概括所得为下面的形式：

$$\mathrm{Zero}(PS) = \mathrm{Zero}(CS/J) + \mathrm{Sum}'\mathrm{Zero}(PS'_k) + \mathrm{Sum}''\mathrm{Zero}(PS''_l),$$

式中出现的符号解释如下。

（CS）是（PS）的强型或弱型的特征组或半特征组。

(PS'_k) 是将多项式 I'_k 添入 (PS) 后的多项式组，这里 I'_k 与 SuM′ 跑遍多项式组 (IS) 中的各个非常数与不重复的多项式，而 (IS) 是 (CS) 中各多项式的初式的集体，在格式 (SCH) 中采用 (BR) 一式时，更包含了历次 (BS_i) 中多项式的各个初式。(PS''_l) 则是将多项式 F''_l 添入 (PS) 后的多项式组，其中 F''_l 与 Sum″跑遍多项式组 (FS) 中的各个非常数与不重复的多项式，而 (FS) 是在格式 (SCH) 中诸 (BS_i) 的各个多项式系数的最大公约式。至于 J 则是这些 I'_k 与 F''_l 的乘积.

整个计算过程，则可概述如下。

设所给多项式组为 (PS)，这里 $(PS)=0$ 为需要求解的方程式组.

依据预定的选择按照格式 (SCH) 从 (PS) 定出三个多项式组：(CS)，(IS) 与 (FS)。

如果 (CS) 由一个常数多项式所组成，则 (PS) 无解。否则如果 (CS) 中多项式的个数少于变量 x_i 的个数 n，则在不可约情形时 (PS) 有无穷多解，这时工作即停止进行，见文 [WU4]。

在 (CS) 恰有 n 个多项式时，可依通常方法将 $(CS)=0$ 的解全部求出，并检查其是否使 $J=0$，而将 $J\neq 0$ 者挑出。

其次将每一 (PS'_k) 与 (PS''_l) 作为新的 (PS) 依同法进行计算。

整个进程在有限步后必然停止。

至此已可判断原来所给的方程组 $(PS)=0$ 是否无解或有无穷多解。而在至多有有限多解的情形下给出全部解答而无增无漏。如果需要的只是（视所给问题而定）整数解或实数解，在解 $(CS)=0$ 时即可用只求整数解或实数解的通常方法，例如对实数解时可用我国宋代的解方程法。

三、程序编制

依据上节的原理与方法，我们编制了若干程序，并在一台以 UNIX 为操作系统的高档微型机 Dual83/20 上就多种理论与应用上的问题进行试算。这些程序，包括方法说明以及实例运算过程，都录制在一片 DYSAN8″软磁盘上，形成一个软件包，总称之为〈解方程器〉或〈SOLUER〉.

实际应用中出现的多项式一般是实系数的，但实系数通过逼近、通分等手段在机器中易于改为用整系数来代替。另一面，不少理论上的问题（例如定理证明）出现的多项式必须是整型的，为了包括各种可能的应用，我们的程序都以整系数多项式的方式来编写。这样每一多项式都相应有 4 个整型的数组来作矩阵形式的表示：*ind*，*ivar*，*iexp*，*icof*，其意义见前节。如有需要，将程序中的系数数组 *icof* 由整型改为实型，是容易的。

程序共有 3 个，列表如下。

由于多项式在运算时须用矩阵形式，而作为原始数据输入或作为结果观察分析时有时又以通常的文字形式为便，故上表中的两程序 1，3 即作为这两种形式互相转换之用。

使用这些程序时，先将原始数据多项式组（*PS*）以通常文字形式存放在文件‘*polgms*’中，然后用程序 1 = *solv-pmat. for* 转变成文件‘*matrices*’中的矩阵形式。其次将‘*matrices*’中的多项式组复制并存放在文件‘*polset*’中，再用程序 2 = *solv-chrst. for* 获得文件‘*char-set*’中的（半）特征组（*CS*），文件‘*init-fctr-set*’中的初式组（*IS*）与公因子组（*FS*），以及记录具体计算过程的文件‘*procedure*’。根据需要可用程序 3 = *solv-mpol*。*fo N* 将存放在‘*cher-set*’

与‘*init－fctr-set*’中的多项式组（*CS*），（*IS*）与（*FS*）等转变成通常文字形式。依据所得（*CS*）可以判断原方程组（*PS*）＝0是否无解，有无穷解，或否，并确定是否应将（*IS*），（*FS*）中的多项式逐个添入（*PS*）而再用以上程序继续进行。

编号	程　序	输　入	输　出	作　用
1	*solv-pmat.for*	以通常形式存放在通道＃3＝‘*polyms*’中的多项式组	以矩阵形式存放在通道＃4＝‘*matrices*’中的多项式组	将一通常形式的多项式组变成矩阵形式
2	*solv-chrst.for*	以矩阵形式存放在通道＃3＝‘*polset*’中的多项式组	以矩阵形式存放在通道＃4＝‘*char-set*’中的特征组或半特征组(*CS*)，在通道＃5＝‘*init-fctr-set*’中的(*IS*)与(*FS*)，以及通道＃6＝‘*procedure*’中的具体计算过程	确定所给多项式组‘*polset*’＝(*PS*)的(*CS*)，(*IS*)，与(*FS*)又将具体计算过程记录备查。
3	*solv-mpol.for*	以矩阵形式存放在通道＃3＝‘*matrices*’中的多项式组	以通常形式存放在通道＃4＝‘*polyms*’中的多项式组	将一矩阵形式的多项式组变成通常形式

在每次所得的（半）特征组（*CS*）中多项式的个数恰等于变量 x_i 的个数 n 时，即可逐一依次解出 x_1，x_2，…，x_n 以获得全部解答，并挑出其中不使任一初式与因式＝0的部分。这需要另一个解方程的程序，由于这样的程序在通常的软件包中往往有现成的可用，而且有多种不同的方法可供选用，因此在我们的软件包〈解方程器〉中不再列入。

在程序运行中出现的各种文件，可以以一定的格式作为永久性的文件保存起来。在［WU5］所举的一些实例中，我们

采用了下面的方式。

文件 * * * * * . pol，存放通常文字形式的原始数据多项式组（PS），即程序 1 中的 ‘polyms’。

文件 * * * * * . mat，存放矩阵形式的原始数据多项式组（PS），即程序 1 中的 ‘matrices’ 或程序 2 中的 ‘*polset*’。

文件 * * * * * . *chr* 与 * * * * * . *inf*，即程序 2 所得的 ‘*char-set*’ 与 ‘*init-fatr-set*’。

文件 * * * * * . *prc*，即程序 2 所得文件 ‘*procedure*’.

三个程序中，程序 2 = *solv-chrst*. *for* 是最主要也是最繁难的一个。为此须作较详的说明如下。

首先，由于计算方式以及打印方式有多种不同的方案可供选择，因此我们引入一些控制变量，其值由使用者依据他的预定要求在键盘上输入，这些控制变量有：

iwt——确定打印部件，=0 指显示，=1 指打印机，=k 指文件 # k （$k>1$）。

iqt——确定多项式的打印方式，<0 指不打印，=0 指打印成通常文字形式，>0 指打印成某种容易阅读的矩阵形式。

ilnp——确定在 *iqt* = 0 时打印出多项式的限制行数。

ips——确定在相应处是否暂停。

ired——确定所需（半）特征组是弱型的（=1）还是强型的（=2）。

init——确定是（〈〉0）否（=0）将初式存入相应文件。

ifctr——确定是（〈〉0）否（=0）须将每一基组中各多项式系数的公因子除去。

在已编成的软件包中，我们并没有考虑一般的系数公因子，而只考虑了单项式的系数公因子。因之在文件 # 5 = ‘*init-fctr-set*’ 中的公因子组，不是一个多项式组，而是一个一维数组 *jfctr*，其个数等于 x_i 的虚拟个数。如果 x_i 曾作为

系数的公因子约去，则 $jfctr\ (i)=1$，否则 $=0$。

在大多数的问题中，是否应约去系数公因子，是可以任意的，可视是否需要将计算过程有较大简化而定，但在某些问题中，系数公因子必须约去，又在另一些问题中，系数公因子又不允许约去，须视问题的性质而定。因而对于控制变量 *ifctr* 在盘面输入时应作何种选择，使用者须事先有一估计。

其他控制变量的输入也须预有计划。

在程序 2 的运算过程中有两套多项式组反复出现，其一相当于格式（SCH）中的诸多项式组（PS_i），在程序中用 a 表示。另一相当于（SCH）中的诸基组（BS_i），在程序中用 b 表示。因此，与 a 相应有数组 *iaind*，*iavar*，*iaexp* 与 *iacof*，程序中又把同一多项式组中的诸多项式 a 合在一起，相应的数组各记为 *jaind*，*javar*，*jaexp* 与 *jacof*，对于多项式 b 同样也有数组 *ibind*，*jbind* 等。

在程序 1 与 3 中，所出现的多项式并不经过任何变化，因而诸 *iexp* 都用二维数组表示，但在程序 2 中，诸多项式在不断运算变化之中，其出现的变量个数时多时少，项数也时多时少。为了减少存储量，这些多项式的 *iaexp*，*ibexp* 等都改用了一维数组。因而，多项式 a 的第 i 项与 *iavar* 中第 j 个数相当的变量的指数将占据着 *iaexp* 中第 $(i-1)*iva+j$ 个位置，这里 *iavar* 是 a 中实际出现的变量 x 的个数。余类推。

程序 2 的运算过程大致如下。

第 1 步　从键盘上读入各个控制变量。

第 2 步　从文件 #3＝‘*polset*’读入原始数据，作为诸多项式 a。

第 3 步　从诸 a 所成多项式组中选取一部分作为多项式 b。相当于前面 a 多项式组的基组，诸 b 的个数记为 *inb*，只

将诸 b 存入文件 #4 = '*char-set*'。这通过子程序 *baset* 来完成。

第 4 步　依据控制变量 *init* 与 *ifctr* 确定是否须定出诸多项式的初式与系数公因子并存入文件 #5 = '*init-fctr-set*' 中。这由子程序 *inset* 与 *factr* 来完成。

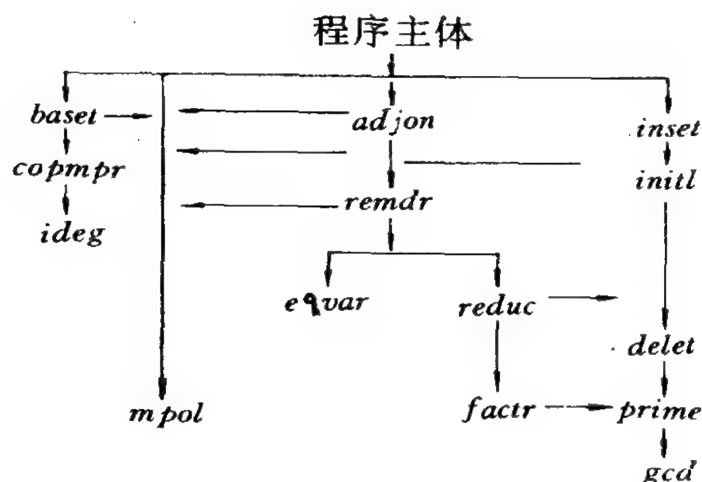


图 1

第 5 步　将诸 a 中不在 b 多项式组者对 b 逐个求余式，将余式不为 0 者添入诸 b。命所得 b 的个数为 *inb*1。

如果 *inb*1 = *inb*，则第 3 步所得 b 多项式组已是所求的（半）特征组而程序中止。否则进入下一步。

这一步由子程序 *adjon* 与 *remdr* 来完成。

第 6 步　将 b 改作为 a，返回第 3 步并继续进行。

在第 5 步中，须不断将某一 a 多项式对某一 b 多项式相除以求余式，而 a 与 b 中实际出现的变量一般并不相同。为此须将两组变量合并为一并相应改变数组 *iaexp* 与 *ibexp* 等再行求余。这些通过子程序 *eqvar* 与 *reduc* 来进行。

各子程序间的相互关系，略如图 1 所示，其中箭头指 *call*。

四、总　　结

作者已将以上软件程序应用之于理论上与实际中经常出现的多种类型的问题，其情况见另文［WU5］。从这些实践得知，这些程序有相当高的效率，但也有许多潜力可挖，上节所说约去多项式系数的公因子尚未写入以上程序中，即是有待发挥的潜力之一。在使用部门根据使用经验提出各种看法意见后，将再统一考虑作软件包的改进型。实践是检验真理的唯一标准，一切当视实践效果如何而定。

参考资料

[R1] J. F. Ritt, Differential equations from the algebraic standpoint, *Amer*, *Math*, S0*c*, 1932.

[R2] J. F. Ritt, Differential algelra, *Amer*, *Math*. S0*c*., 1950.

[WU1] 吴文俊，初等几何判定问题与机械化证明，中国科学（1977），507—516，或即

Wu Wen-tsün, On the decision problem and the mechanization of theorem proving in elementary geometry, *Scentuia Sinica*, 21 (1978), 159—172 以及 Automated theorem proving: After 25 years *Contemporary Mathematics*, 29 (1984), 213—234.

[WU2] Wu Weu-tsün, Basic priuciples of mechanical theorem proving in clementary geometies, *J*. *Sys*, *sci* & *Math*. *Scis*., 4 (1984), 207—235.

[WU3] 吴文俊，几何定理机器证明的基本原理（初等几何部分），科学出版社，北京，1984。

[WU4] 吴文俊，关于代数方程组的零点——Ritt 原理的一个应用，科

学通报（1985）881—3。
[WU5] 吴文俊，〈解方程器〉或〈SOLVER〉软件系统应用实例，见本刊下期。

解方程器或 SOLVER 软件系统应用举例*

引　　言

前文 [WU6] 介绍了解方程器或 SOLVER 软件系统的基本原理与方法，以及程序的编制与使用。本文将介绍使用这些程序对理论上与应用上提出的若干实例在一台微型机Dual83/20上实施的经过情况。在这些实例中，有的是纯理论性的，包括以三角形面积与四面体体积为说明的未知公式探求方法，以及受到 Gauss 注意过的一条著名初等几何定理的证明。在应用上，则除一般的解高次联立方程组外，还举出求特征值、控制论中的反馈稳定与机械设计中出现的某些实例，读者可以将我们的方法与已有的各种方法作一比较，我们的方法是带有普遍性的，最后都归结为高次联立方程组的求解，其应用范围遍处皆有，读者举一反三，尽可能在其自己的工作范围内找到合适的应用课题，并据此以判断〈解方程器〉的效率。

现再提出两点作为补充说明。

第一，我们的方法基本上是线性联立方程组化成三角阵式的方法及其在高次联立方程组的推广，这在国外通称为 Gauss 消去法，有时也称为 Gauss-Seidel 方法。但 Seidel 事实上与此无关，而 Gauss 引入此法是在 1826 年的关于天文观察计算的

---

* 本文摘自《数学的实践与认识》，1986 年第 3 期。

论文中，其中只有带对称性线性联立方程组的实例见［G1］。实质上这种消去法早已见于西汉早期的我国古代经典著作〈九章算术〉，而且方法上还完整得多。在高次联立方程组上的推广，也早已在宋元时代秦九韶、李冶、朱世杰等的传世经典著作中见其雏形。我们现在的方法。其发现过程受到这些著作中所蕴含的思想、方法、体系等的影响与启发，是有决定性的。因之，作者一直自己认为，目前的工作以及过去的机器证明工作，都是我国古代数学传统的直接继承。此其一。

第二，在作者对于本文中提出的一些实例以及其他一些实例使用〈解方程器〉具体进行计算时，发现目前的程序还有许多不足之处或其潜力尚未得到充分发挥。例如，由于在程序中我们没有考虑到将多项式中系数的公因子（文献中流行名称是 Content）约去，因而所得的多项式次数往往较应有者为高，也因而增加了不必要的计算量与存储量，例如参阅下面的例子。此外也可能还有许多弱点未被作者所发现或考虑。总之，实践是检验真理的唯一标准。根据各界在使用〈解方程器〉解决理论上与应用上各种具体问题时的实践经验，考其得失与有效或不足之处，当在将来再在总结经验的基础上作出相应改进。此其二。

例 1　方程组求解。

与解线性联立方程组类似，解高次联立方程组也有两种一般的方法；叠代逼近法与消去法。前者对于叠代收敛性在理论上似乎还没有一般的准则，后者就作者所知，有两种较通用的方法，为法国 Lazard 与奥地利 Buchberger 所创立，例如见 [LZ1]，［BU1］，现就［BU1］中 Buchberger 所举一例说明如下。

Buchberger 考虑了方程组 $A_i=0$，$i=1, 2, 3$。其中 A_i 各为多项式

$$A_1 = 2x_3^2 - x_1^2 - x_2^2,$$
$$A_2 = x_1x_3 - 2x_3 + x_1x_2,$$
$$A_3 = x_1^2 - x_2。$$

按照 Lazard 的方法，需要引进 4 个新变量 u_0，u_1，u_2，u_3，并考虑一个由 u_i 等所构成的矩阵 Φ，其行数与列数各为 35 与 50。从 Φ 算出一个 35 行的非 0 行列式，其本身为 u_i 的多项式。将此多项式分解为 u_i 线性式的因子乘积，即得出原方程组包括在无穷远处的解答。显然这一方法不是切实可行的。

Buchberger 所用方法在于先将诸多项式构成一个理想 $a=(A_1, A_2, A_3)$，由此得出一组所谓 Gröbner 基，在本例中由 6 个多项式组成

$$G_1 = x_3^2 - \frac{1}{2}x_2^2 - \frac{1}{2}x_2,$$
$$G_2 = x_1x_3 - 2x_3 + x_1x_2,$$
$$G_3 = x_1^2 - x_2,$$
$$G_4 = x_1x_2^2 + 7x_1x_2 + 2x_2 + 6x_2^2 - 16x_3,$$
$$G_5 = x_2x_3 + x_2^2 + 2x_1x_2 - 4x_3,$$
$$G_6 = x_2^3 - 12x_2 - 29x_2^2 + 64x_3 - 24x_1x_2。$$

由此得出商代数 $R[x_1, x_2, x_3]/a$ 的一组基

$$1, \overline{x_1}, \overline{x_2}, \overline{x_3}, \overline{x_1x_2}, \overline{x_2^2}。$$

从 Gröbner 基可定出这些基之间在商代数中的乘法关系，例如，

$$\overline{x_2^2} * \overline{x_3} = -18\,\overline{x_1x_2} + 48\overline{x_3} - 8\overline{x_2} - 21\overline{x_2^2}。$$

据这些乘法关系再定出 $\overline{x_1}$，$\overline{x_2}$ 等所满足的一些方程，$p_1(\overline{x_1})=0, p_2(\overline{x_1},\overline{x_2})=0, \cdots$。解出这些方程即得到原方程的解答，但不能保证并无增根。在理论上似乎也还没有解决如何求出全

部解答而无增无漏的问题。

我们的方法如下：

将所给多项式 A_1，A_2，A_3 所构成的多项式组（PS）作为原始数据输入解方程器，得出历次基组如次。

第一次基组（$BS1$）包括两个多项式：

$$B1_1 = x_1^2 - x_2,$$

$$B1_2 = x_1x_3 - 2x_3 + x_1x_2。$$

第二次基组（$BS2$）包括 3 个多项式：

$$B2_1 = x_1^4 - 5x_1^2 + 4x_1 - 4 + 4x_1^3,$$

$$B2_2 = x_1^2 - x_2,$$

$$B2_3 = x_1x_3 - 2x_2 + x_1x_2。$$

这一基组已是所求的半特征组（CS），其中 3 个多项式是 $C_i = B2_i$，$i = 1$，2，3，在计算过程中出现的非常数初式只有一个 $I = x_1 - 2$。约去过的因子也只有一个 $F = x_1$。

依据一般原理（见前一文［WU6］），应有

$\mathrm{Zero}(PS) = \mathrm{Zero}(CS/IF) + \mathrm{Zero}(PS1) + \mathrm{Zero}(PS2)$，这里 $(PS1)$，$(PS2)$ 各为将 I 与 F 添入 (PS) 后所得的多项式组。

方程组 $(CS) = 0$ 已是标准的三角化形式，可以逐个对 x_1，x_2，x_3 依次求解，得到 4 个解答，但其中一个使 $I = 0$，因而必须除去。余下 3 个解答都是 $(PS) = 0$ 的真正解答。

其次试考虑 $\mathrm{Zero}(PS1)$，这里 $(PS1)$ 中除前面的 A_1，A_2，A_3 外，又加入了 $A_4 = I$：

$$A_4 = x_1 - 2。$$

与前同法得到半特征组由一个常数多项式所组成，因而是矛盾组而无解答。至于（$PS2$）无解答更甚显然。因而所给方程的全部解答即是前面所得 3 个而无增无漏。

所有计算过程都有打印记录。这些记录以及其他有关文件

都以 Buchberger 与 Buchbergl 命名而作为软件包的一部分存入解方程器中。

要求得无穷远处的解答是容易的。因为这些无穷远处的解答相当于只保留所给多项式的最高次部分所得方程组的解答，即考虑以下方程组的解答：

$$A_{1\infty}=2x_3^2-x_1^2-x_2^2,$$

$$A_{2\infty}=x_1x_3+x_1x_2,$$

$$A_{3\infty}=x_1^2$$

这比所给方程式组显然更容易求得所有解答。

例 2 求特征值与特征根。

矩阵求特征值与特征根是实际应用中广泛出现的问题。现在通行的办法往往依据问题的不同类型（实或复，对称或非对称，广义或非广义形式等等）有着种种不同的处理，极为复杂，例如参阅［RW1］与［郭 1］等书。我们的方法则对任意类型都可作统一处理。现据［郭 1］书中 § 13.1.11 中的一例作为说明如下。

该例要求矩阵方程

$$AX=eX$$

的特征值 e 与特征根 $X=(x_1,\cdots,x_6)$，这里，

$$A=\begin{bmatrix} 5 & 1 & -2 & 0 & -2 & 5 \\ 1 & 6 & -3 & 2 & 0 & 6 \\ -2 & -3 & 8 & -5 & -6 & 0 \\ 0 & 2 & -5 & 5 & 1 & -2 \\ -2 & 0 & -6 & 1 & 6 & -3 \\ 5 & 6 & 0 & -2 & -3 & 8 \end{bmatrix}。$$

我们的方法在于将 e 作为另一未知变量 x_0 引入，于是原矩阵方程变为 7 个未知量 x_0，…，x_6 的方程组（PS）$=0$，

而（PS）由 6 个多项式 A_i，$i=1, 2, \cdots, 6$ 组成，例如

$$A_1=5x_1+x_2-2x_3-2x_5+5x_6-x_0x_1,$$

将（PS）输入解方程器，计算结果第 4 个基组为所求半特征组（CS），由 6 个多项式 C_1，…，C_6 所组成如下。

$$C_1=325x_0^2-277x_0-51x_0^3-1495+2x_0^4,$$

$$C_2=-2300x_1-780x_0x_1+380x_0^2x_1+3335x_2$$
$$+1131x_0x_2-551x_0^2x_2-20x_0^3x_1+29x_0^3x_2,$$

$$C_3=29x_1-18x_2-13x_3+2x_3x_0-6x_1x_0+5x_2x_0,$$

$$C_4=-69x_2+166x_3-145x_4-58x_1+x_3x_0$$
$$+12x_1x_0+30x_4x_0,$$

$$C_5=-2x_1-3x_2+8x_3-5x_4-6x_5-x_3x_0,$$

$$C_6=5x_1+x_2-2x_3-2x_5+5x_6-x_1x_0。$$

由此可以先从 $C_1=0$ 解出 x_0 共有 4 个值可作为可能的特征值，依次代入 $C_2=0$，…，$C_6=0$ 可得相应的 $x_1:x_2:\cdots:x_6$ 即相应可能的特征根。

在计算过程中出现唯一约去的因子为 x_1，非常数的初式有 3 个如下：

$$I_2=3335+1131x_0-551x_0^2+29x_0^3,$$

$$I_3=2x_0-13,$$

$$I_4=6x_0-29。$$

必须除去 $C1=0$ 中使某一因子 $F_1=x_1$ 或初式 $I_i=0$ 的解，这可以用求 F_1，I_i 与 C_i 有否公因子的方法来检查。在本例，可以检查得知 C_1 与 I_3 有公因子 $2x_0-13$，且 $C_1/I_3=x_0^3-19x_0^2+39x_0+115$。其次须将 F_1，I_1，I_2，I_3 逐一添入原来的多项式组（PS）并如前处理，但并不能获得其他解答。因而特征解即为

$$C1/I3=0$$

所给出的 3 个，特征根也有相应的 3 组，以 $x_1:x_2:\cdots:x_6$ 的形式给出。读者可以与［郭 1］书中给出特征值的解答对照：

$$\lambda_1=\lambda_2=-0.1598734292_{10}1,$$

$$\lambda_3=\lambda_4=0.4455989638_{10}1,$$

$$\lambda_5=\lambda_6=0.1614274469_{10}2。$$

作者还以同法计算了该书 § 13.1.11 的一个 $AX=eBX$ 的例子。方法完全与上面一样，不过最后得一 x_0 的 16 次方程，须同上面那样把不需要的解答除去。次数出现过高是因为在程序中未考虑约去一般的公因子之故，当于以后加以改进。

例 3　自动控制问题。

控制论中不少问题，如极点配置，反馈稳定等，都自然地导致高次联立方程组的求解问题。下例取自［AS1］一文，读者可与它对照。

考虑一个线性系统

$$\dot{x}=\begin{bmatrix}0 & 1 & 0\\ 0 & 0 & 1\\ 0 & 13 & 0\end{bmatrix}x+\begin{bmatrix}0\\ 0\\ 1\end{bmatrix}u,$$

$$y=\begin{bmatrix}0 & 5 & -1\\ -1 & -1 & 0\end{bmatrix}x。$$

现在要求确定一个反馈控制

$$u=ky,\quad k=[v\ w],$$

使系统达到稳定，依据控制论一般原理，应选取 v，w 实值，使以下诸不等式成立：

$$p_1(v,w)=w(v-1)-v(5v+13)>0,$$

$$p_2(v,w)=v>0,$$

$$p_3(v,w)=w>0。$$

这些不等式相当于有实值 t_i，使

$$t_i^2 f_i(v, w) - 1^1 = 0,\ i = 1, 2, 3。$$

这样须求 5 个未知数 t_i 与 v， w 满足 3 个方程。

在无数可供选择的不同方案中，［AS1］的作者提出要求某一目标函数

$$J_c = \sum_{j=1}^{3}(x_j - \delta_j)^2 + \sum_{j=1}^{3}(t_i - r_i)^2$$

达到极小，并指出必有 r_i, δ_j 能使满足要求的解答只有有限多个。特别对上例将导致下面的一个方程组：

$$t_1^2(-5v^2 + wv - 13v - w) - 1 = 0,$$

$$t_2^2 v - 1 = 0,$$

$$t_3^2 w - 1 = 0,$$

$$(\delta_1 - v) - \frac{1}{2}t_1^3(r_1 - t_1)(-10v + w - 13)$$

$$- \frac{1}{2}t_2^3(r_2 - t_2) = 0,$$

$$(\delta_2 - w) - \frac{1}{2}t_1^3(r_1 - t_1)(v - 1)$$

$$- \frac{1}{2}t_3^3(r_3 - t_3) = 0。$$

[AS1]一文选取 $r_i = \delta_j = 0$，消去 t_i 后得两个方程。

$$g_1(v, w) = a_2(v)w^2 + a_1(v)w + a_0(v) = 0,$$

$$g_2(v, w) = b_5(v)w^5 + b_4(v)w^4 + b_3(v)w^3$$

$$+ b_2(v)w^2 + b_1(v)w + b_0(v) = 0,$$

其中 a_i, b_j 都是以小数为系数的 v 的多项式。这一步似乎由手算作出，某些系数有明显的错误。[AS1]作者用结式方法消去 w 而得一 v 的 31 次多项式方程，系数是复杂的小数，由此得所需的解。

我们也用解方程器进行了试算。先将诸变量($v, w, 1/t_1, 1/t_2, 1/t_3$)改变为($x_1, x_2, \cdots, x_5$),得多项式组($PS$):

$$A_1 = -5x_1^2 + x_1x_2 - 13x_1 - x_2 - x_3^2,$$

$$A_2 = x_1 - x_4^2,$$

$$A_3 = x_2 - x_5^2,$$

$$A_4 = -2x_3^4x_4^4x_1 - 10x_4^4x_1 + x_4^4x_2 - 13x_4^4 + x_3^4,$$

$$A_5 = -2x_3^4x_5^4x_2 + x_5^4x_1 - x_5^4 + x_3^4。$$

所求者为 Zero ($PS/x_3x_4x_5$)中的实数解, 而 x_3, x_4, x_5 都不应为 0, 计算至第 4 个基组即所求半特征组 (CS), 其第一个多项式 C_1 是 x_1, 即 v 的一个 47 次多项式。这里所得的多项式次数较 [AS1] 一文者为高, 其原因是在计算过程中某些多项式的系数公因子未曾约去所致, 而这种公因子的必然出现是在理论上可以证明的。这在今后的程序中应当加以考虑改写, 据估计改写后计算将较 [AS1] 者为简单。而且, 国外所使用结式方法一般说来是难以实现的, 而我们的方法适用范围似更广。

例 4 机构综合设计。

机构的综合设计, 往往导致高次联立方程组的求解问题, 下面是一个摇杆滑块的设计问题, 许多问题都可仿此求解。

设要求设计一个摇杆滑块装置 O_aAB, 其摇杆的转动中心 O_a 固定在原点, 滑块在一平行于横轴的一个槽上滑动, 设计要求是对于 3 个已知的摇杆 O_aA 对横轴的倾斜角度 θ_i, 滑块 B 与纵轴有已知的距离 S_i, $i=1, 2, 3$。

通常的解决办法有两种, 一种是几何作图法, 另一是代数方法, 前者难以精确且作图复杂, 后者会引入许多三角函数也有颇大的计算量。我们的方法则可避免使用三角函数, 而在多项式的非数值运算间进行, 至最后再引入原始数值数据而对方

程求数值解，方法是一般性的，对于多连杆或其他更复杂的机构设计也同样适用

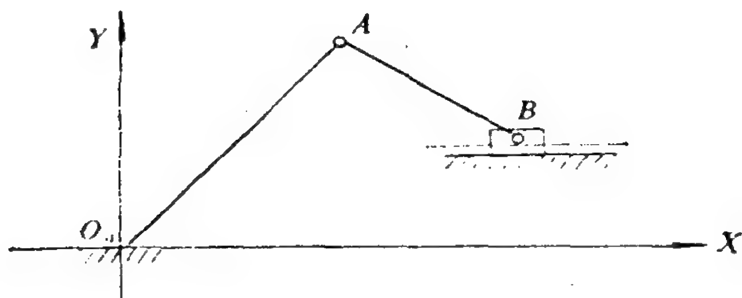


图 1

对于上举简单实例来说，我们引入变量

$$x_{11}=\tan\frac{1}{2}\theta_1=\tan\frac{1}{2}\angle A_1O_aX,$$

$$x_{12}=\tan\frac{1}{2}\theta_2=\tan\frac{1}{2}\angle A_2O_aX,$$

$$x_{13}=\tan\frac{1}{2}\theta_3=\tan\frac{1}{2}\angle A_3O_aX。$$

这里 $O_aA_iB_i$ 是装置的三个不同位置，于是 $\sin\theta_i$，$\cos\theta_i$ 都是 x_{11}，x_{12}，x_{13}的有理函数，取 A_i，B_i 不同位置的坐标各为

$$A_1=(x_{20},x_{21}),A_2=(x_{22},x_{23}),A_3=(x_{24},x_{25}),$$

$$B_1=(x_{14},x_{29}),B_2=(x_{15},x_{29}),B_3=(x_{16},x_{29})。$$

其中 x_{14}，x_{15}，x_{16}即滑块 B 对纵轴的已知距离 s_1，s_2，s_3，但我们不必将 s_i 的已知具体数值代入，而将 x_{14}，x_{15}，x_{16}作为不定量来处理，对 x_{11}，x_{12}，x_{13}亦然。

其次取摇杆 O_aA 与连杆 AB 的长度为x_{18}与 x_{30}。于是设计的要求是从 x_{11}，…，x_{16}的具体所给数值求得 O_aA，AB 杆的长度 x_{18}，x_{30}以及滑块离开横轴的高度 x_{29}。

这里牵涉到的变量有 15 个：

$$x_{11}, x_{12}, x_{13}, x_{14}, x_{15}, x_{16}, x_{18},$$

$$x_{20}, x_{21}, x_{22}, x_{23}, x_{24}, x_{25}, x_{29}, x_{30}。$$

其中 x_1，…，x_{10}，x_{17}，x_{19}，x_{26}，…，x_{28}并不出现。这并无任何关系，重要的是这些 x 出现的次序。我们把实际已知的变量排在最前（x_{11}，…，x_{16}），其次是 x_{18}，这样可首先求出 O_aA 的长度 x_{18}，再依次求出其他诸 x。

依题意立方程，得相应的多项式组（PS）并应用解方程器程序进行计算时，计算过程中将迅速出现几千项的多项式，其原因是我们目前所编程序并没有考虑到将必然会出现的某些多项式公共因子（在本例是 $1+x_{11}^2$，$1+x_{12}^2$ 与 $1+x_{13}^2$）约去之故。为避免将程序作较大更动而使项数过多的多项式不致出现，我们采取了下面的权宜办法。为此引入新变量 x_1，x_2，x_3，使

$$x_1=1+x_{11}^2,\quad x_2=1+x_{12}^2,\quad x_3=1+x_{13}^2。$$

于是与依题意立方程所得相应的多项式组（PS）系由以下 12 个多项式所组成：

$$A_1 = 1 + x_{11}^2 - x_1',$$

$$A_2 = 1 + x_{12}^2 - x_2,$$

$$A_3 = 1 + x_{13}^2 - x_3,$$

$$A_4 = x_1 x_{20} - x_{18} + x_{11}^2 x_{18},$$

$$A_5 = x_1 x_{21} - 2x_{11} x_{18},$$

$$A_6 = x_2 x_{22} - x_{18} + x_{12}^2 x_{18},$$

$$A_7 = x_2 x_{23} - 2x_{12} x_{18},$$

$$A_8 = x_3 x_{24} - x_{18} + x_{13}^2 x_{18},$$

$$A_9 = x_3 x_{25} - 2x_{13} x_{18},$$

$$A_{10} = (x_{20} - x_{14})^2 + (x_{21} - x_{29})^2 - x_{30}^2,$$

$$A_{11} = (x_{22} - x_{15})^2 + (x_{23} - x_{29})^2 - x_{30}^2,$$

$$A_{12} = (x_{24} - x_{16})^2 + (x_{25} - x_{29})^2 - x_{30}^2。$$

用解方程器计算至第 4 个基组即为所求半特征组(CS),其中前三个多项式 C_1, C_2, C_3 即原来的 A_1, A_2, A_3,而第 4 个多项式 C_4 是一只有 18 项的 x_{18}的一次多项式:

$$C_4 = \begin{cases} 2x_{18}\begin{bmatrix} x_{14}(x_1 - 2)(x_2 x_{13} - x_3 x_{12}) \\ + x_{15}(x_2 - 2)(x_3 x_{11} - x_1 x_{13}) \\ + x_{16}(x_3 - 2)(x_1 x_{12} - x_2 x_{11}) \end{bmatrix} \\ - \begin{bmatrix} x_1 x_{14}^2 (x_2 x_{13} - x_3 x_{12}) \\ + x_2 x_{15}^2 (x_3 x_{11} - x_1 x_{13}) \\ + x_3 x_{16}^2 (x_1 x_{12} - x_2 x_{11}) \end{bmatrix} \end{cases}$$

最后的两个多项式 C_{11}是 x_{29}的一次多项式,项数为 103,而 C_{12}即原来的多项式 A_{10}。

由于 x_1,x_2,x_3,x_{11},…,x_{16}都是已知的,将这些数代入 C_i 后,即可从 $C_1 = 0$ 解出 x_{18},又逐个代入 $C_2 = 0$,…,$C_{12} = 0$ 以最后解出 x_{29},x_{30},即得所求的一个设计方案。

在求（CS）的计算过程中所出现的初式与约去的公共因子为

$$x_1,\ x_2,\ x_3,\ x_{18},\ x_1x_{12}-x_2x_{11}。$$

由于 x_1，x_2，x_3，x_{11}，x_{12} 都是已知数，我们可假设 x_1，x_2，x_3，$x_1x_{12}-x_2x_{11}$ 都不为 0。事实上 $x_1=1+x_{11}^2$ 等也不可能为 0。至于 $x_{18}=0$ 的不合理是显然的，因此一般说来已不必再将这些初式或公共因子添入（PS）的进一步考虑（PSk'）与（PSl''），这时只有一个装置可供采用，由于

$$\begin{aligned}x_1x_{12}-x_2x_{11}&=(1+x_{11}^2)x_{12}-(1+x_{12}^2)x_{11}\\&=(x_{12}-x_{11})(1-x_{11}x_{12})\\&=\left(\tan\frac{1}{2}\theta_2-\tan\frac{1}{2}\theta_1\right)\\&\quad\left(1-\tan\frac{1}{2}\theta_1\tan\frac{1}{2}\theta_2\right),\end{aligned}$$

故 $x_1x_{12}-x_2x_{11}=0$ 时，有 $\theta_1\pm\theta_2=\pi$（设 $\theta_1>\theta_2>0$）。如果原始数据中的 θ_i 符合这种特殊情况，则也可将 $x_1x_{12}-x_2x_{11}$ 添入（PS）并用解方程器求得相应的特殊装置。

例 5 面积与体积。

解方程器不仅可用于各种实际应用问题，也可用于纯理论的探索，本例关于公式或关系的探求以及下例关于定理的证明与发明都属此类。现就三角形的面积与四面体的体积公式作一说明。

设三角形的三边之长为 a，b，c，面积为△，则西方有所谓海伦公式

$$\triangle=\sqrt{s\ (s-a)\ (s-b)\ (s-c)},$$

其中 $s=\frac{1}{2}\ (a+b+c)$。

在我国，则秦九韶在〈数书九章〉（1247）中，给出了公式

$$\triangle=\sqrt{\frac{1}{4}\left[\text{小}^2\cdot\text{大}^2-\left(\frac{\text{大}^2+\text{小}^2-\text{中}^2}{2}\right)^2\right]},$$

其中大、中、小为三边之长，即 a，b，c。现存所谓海伦公式的证明是极其迂迴曲折难以知其思路的，这无异于说明海伦其人海伦其证都来历不明，十分可疑。另一面作者曾依据我国古代经典《九章算术》与刘徽《九章注》的传统方法补上了一个秦九韶公式的证明，最后自然地导致秦九韶公式，而无雕琢穿凿之嫌，参阅［WU7］。

不论是海伦的证明或是作者对于秦九韶公式的证明，都须要比较巧妙而特殊的思考，特别是前者。如果事先不知道所要证明的公式，更不知从何下手。只要想像一下如何从四面体 6 个棱的长度求得其体积的问题就可知其困难程度。然而解方程器却提供了一种方法，可以不经过艰苦的思考，轻易地得出未知的公式并包括所需的证明。下面以三角形面积公式的探求来加以说明，这个方法是极为一般的，可以适用于多种纯数学理论中出现的问题。

为简化计算起见，将取三角形 a_0，a_1，a_2 的一个顶点 a_0 在原点，a_0a_1 为横轴，取变量 x_i 如下：

长 $a_0a_1=x_1$，长 $a_0a_2=x_2$，长 $a_1a_2=x_3$，

面积 $a_0a_1a_2=x_4$，

$a_1=(x_5,0), a_2=(x_6,x_7)$。

这里 x_i 取法的先后次序极其重要。我们把三个边的已知长度作为最先出现的 x_1，x_2，x_3，其次就是所求的面积 x_4，最后 a_1，a_2 的坐标用 x_5，x_6，x_7 表示，是无足轻重的。计算所得的（半）特征组将首先给出 x_4 与 x_1，x_2，x_3 的关系，即将是所求的面积公式。

今依题意立方程，得由 4 个多项式组成的多项式组，各相当于面积与边长的表示式：

$$A_1 = 2x_4 - x_5x_7,$$
$$A_2 = x_5 - x_1,$$
$$A_3 = x_2^2 - x_6^2 - x_7^2,$$
$$A_4 = x_3^2 - x_6^2 - x_5^2 + 2x_6x_5 - x_7^2。$$

用解方程器计算至第 3 个基组即为所求半特征组（CS），其第一个多项式为

$$C_1 = 2x_1^2x_3^2 - x_3^4 + 2x_2^2x_3^2 - x_2^4 - x_1^4 + 2x_1^2x_2^2 - 16x_4^2。$$

依据一般理论（见 [WU6]），原多项式组的零点也是（CS）的零点，即每一符合假设的几何图象（在本例是边长为 x_1，x_2，x_3 的三角形）。必然满足公式 $C_1 = 0$ 等。这就给出了事先并不知道而在计算过程中自然得出的公式 $C_1 = 0$，即秦九韶-海伦的面积公式，计算过程同时也就是这一公式的证明过程。

同样的方法也给出了四面体体积用六稜之长来表达的公式：设四面体 $a_0a_1a_2a_3$ 的稜长为 $a_0a_i = x_i a_i a_j = x_{i+j+1}$，$i$，$j = 1$，2，3，而体积为 x_7，则计算给出

$$\begin{aligned}144x_7^2 = & -(x_1^2x_6^4 + x_1^4x_6^2 + x_2^2x_5^4 + x_2^4x_5^2 + x_3^2x_4^4 + x_3^4x_4^2) \\ & + x_2^2x_5^2(x_1^2 + x_3^2 + x_4^2 + x_6^2) \\ & + x_1^2x_6^2(x_2^2 + x_3^2 + x_4^2 + x_5^2) \\ & + x_3^2x_4^2(x_1^2 + x_2^2 + x_5^2 + x_6^2) \\ & - (x_1^2x_2^2x_4^2 + x_1^2x_3^2x_5^2 + x_2^2x_3^2x_6^2 + x_4^2x_5^2x_6^2)。\end{aligned}$$

顺便指出，Cayley 曾利用行列式的巧妙计算得出上面的公式，但我们的推导不仅不需要巧思，甚至几乎不加思索。而且，同样的方法也可用于求其他类似的公式。例如，在双曲型非欧平面中，三角形的面积是三角形之和与 π 之差，但如何将面积用三边之长表示出来，一般不为人知，用我们的方法也可自然地导出。通常非欧几何要广泛使用三角函数与双曲函数等超越

函数，是完全可以避免的，参阅［WU3］一书。

例6 定理证明。

作者以及周咸青、王东明、胡森等同志都编有专用的机器证明程序，并用以证明并发明（= 发明 + 证明）了许多定理。但如果只用解方程器，虽效果或不如专用程序，但也提供了证明与发明定理的有效手段，其方法与原理犹如下述。将所要求证的定理的假设部分表达为（PS）$=0$ 的形式，这里（PS）是用适当坐标表达下的多项式组。现在引入一新变量 x_n，n 较已出现 x_i 的下指标 i 都大。记所需求证的结论为 $R=0$。作多项式 $x_n - R$ 添入（RS）以得新多项式组（PS'）。对（PS'）用解方程器进行运算，如果最后所得（半）特征组（CS）中有一多项式作形式 $I_n x_n$，这里 I_n 是相应初式，其中只有 x_n 之前的变量 x_i，即可得出结论：在 $I_n \neq 0$ 的这一非退化条件下，定理 $x_n=0$ 或 $R=0$ 成立。

这个方法是一般性的，也可用以发明新定理。现举一具体实例如次。

平面中任意四条直线构成一个所谓完全四边形（假设无平行与三线共点）。四直线的六个交点称为顶点，其中任两直线的交点与另两直线的交点称为一组对顶点。其联线称为对角线。这样的对角线有三。一个著名的定理是说：

完全四边形的三个对角线的三个中点位于一直线上。

我们将称这一直线为完全四边形的中点线。在连杆机构综合设计的几何理论中，中点线有着一定的实际意义。

另一面，Gauss 在一篇 1810 年的论文［G2］中，指出了这一定理。Gauss 论文的目的在求与四直线都相切的椭圆面积的最大值。在解决这一问题的过程中，Gauss 指出与四直线相切的所有二次曲线的中心都在一直线上，特别当二次曲线退化为相应完全四边形的一组对顶点时，其中心即为相应对角线的

中点，因此作为特例，三个对角线的中点位于一直线上，即上面的中点线定理。

Gauss 对这一结果作了一个说明：

这一对任意四边形都成立的一般性质，就我（Gauss 自称）所知还没有人注意过。因此在下面我将给出一个简单的直接证明。

除了 Gauss 在文末给出的简单证明外，也还有其他的多种证明。但在 Gauss 之前是否早已有人发现过这一定理，则有待查考。

用解方程器证明这一定理的方法如次。

设所给四直线为 L_1，…，L_4。为计算简单起见，将其中 L_1，L_2 作为斜坐标轴，命诸交点的坐标为

$$a_0 = L_1 \wedge L_2 = (0,0),\ a_1 = L_1 \wedge L_3 = (x_1, 0),$$
$$a_2 = L_1 \wedge L_4 = (x_2, 0),\ a_3 = L_2 \wedge L_4 = (0, x_3),$$
$$a_4 = L_2 \wedge L_3 = (0, x_4),\ a_5 = L_3 \wedge L_4 = (x_5, x_6)。$$

记对角线 a_0a_5, a_1a_3, a_2a_4 的中点为 m_1, m_2, m_3，并设

$$m_1 = (x_7, x_8),\ m_2 = (x_9, x_{10}),\ m_3 = (x_{11}, x_{12})。$$

于是与假设相应的多项式组由以下 8 个多项式所组成：

$$A_1 = x_1x_4 - x_4x_5 - x_1x_6,$$
$$A_2 = x_2x_3 - x_3x_5 - x_2x_6,$$
$$A_3 = 2x_7 - x_5,$$
$$A_4 = 2x_8 - x_6,$$
$$A_5 = 2x_9 - x_1,$$
$$A_6 = 2x_{10} - x_3,$$
$$A_7 = 2x_{11} - x_2,$$
$$A_8 = 2x_{12} - x_4。$$

需要证明的结论 m_1, m_2, m_3 在一直线上相当于

$$x_8x_{11}+x_7x_{10}+x_9x_{12}-x_8x_9-x_{10}x_{11}-x_7x_{12}=0,$$

其左边也是三角形 $m_1m_2m_3$ 的面积。为此引入变量 x_2。以代表这一三角形面积，并将

$$A_9=2x_{20}-x_8x_{11}-x_7x_{10}-x_9x_{12}+x_8x_9+x_{10}x_{11}+x_7x_{12}$$

添入前一多项式组，我们的目的是在于证明在某些非退化条件下，

$$x_{20}=0。$$

用解方程器进行计算至第 4 个基组为所求半特征组，系由 9 个多项式组成，最后一个为

$$C_9=(-x_1^3x_3^2+2x_1^2x_4x_2x_3-x_1x_4^2x_2^2)x_{20}。$$

在计算过程中出现的诸非常数初式都是因子

$$x_1, x_1x_3-x_2x_4$$

等的乘积。由此知在非退化条件

$$x_1\neq0, x_1x_3-x_2x_4\neq0$$

下 $x_{20}=0$，而定理成立。

这两非退化条件的几何意义是显然的。前者指 $a_1\neq a_0$，即 L_1，L_2，L_3 不经过同一点，后者指 L_3，L_4 不相平行。因而可以得出结论：定理普遍成立。

王东明应用他与胡森同志合编的机器证明程序，发现并证明也就是发明了下面有趣的定理：

王东明-胡森定理 任给 5 条直线，任取其中 4 条所得完全四边形的 5 条中心线必经过同一点。

自然用解方程器而不用专门的证明程序也可以证明（但已不是发明）这一定理。

参 考 文 献

[AS1] B. D. O. Anderson & R. W. Scott, Output feedback stabiliza-

tion—solution by algebraic geometry methods, *Proc. IEEE*, 65 (1977), 849—861.

[BU1] B. Buchberger, Ein algorithnicsches Kriterium für die Lösbarkeit eines algebraischen Gleichungs systems, *aeq. math.*, 4 (1970), 374—383.

[G1] K. F. Gauss, Supplementum theoriae combinationis observationum erroribus minimis obnoxiac (1826), Werke, Bd. 4 (1880), 57—93.

[G2] K. F. Gauss, Bestimmung der grössten Ellipse welche die vier Seiten eines gegebenen Vierecks berührt (1810), Werke, Bd4 (1880), 385—392.

[LZ1] D. Lazard, Systems of algelraic eguations Symb & Alg. Comp., Lect. Notes in Comp. Sei., #72 (1977), 88—94.

[RW1] A. Ralston & H. S. wilf, 徐献瑜等译，数字计算机上用的数学方法，上海科技出版社，1963。

[郭 1] 郭富印等，Fortran 算法汇编（第 3 分册），国防工业出版社，北京，1982。

[WU6] 吴文俊，解方程器或 SOLVER 软件系统概述，数学的实践与认识，2 (1986)。

[WU7] 吴文俊，出入相补原理，中国古代科技成就，中国青年出版社，北京，1978，80—100。

解方程器软件包（EQNS－SOLVER）*

一、引　　言

解方程器软件包（EQNS－SOLVER）的目标是给出整系数多项式方程的实数或复数解。这里我们假定有 n 个未知数 x_1，……，x_n 和 r 个方程

$$
\begin{aligned}
&A_1(x_1, \cdots\cdots, x_n)=0,\\
&A_2(x_1, \cdots\cdots, x_n)=0,\\
&\cdots\cdots\cdots\cdots\cdots\cdots \qquad \text{(EQN-A)}\\
&A_r(x_1, \cdots\cdots, x_n)=0。
\end{aligned}
$$

我们将方程组 EQA-A 左边的多项式组成的集合记为 PS-A，而将 EQN-A 方程组的全部复数解记为 Zero（PS-A），如果给出另一个多项式 G，则对应于 $G<>0$ 这个条件的 EQN-A 全部复数解记为 Zero（PS-A/G）。

解方程器软件包将精确地回答 Zero（PS-A）这个集合是否是一个空集、无限集、或者是有限集。换言之，给出这个 EQN-A 方程系统是否无解、无穷个解或者有有限个解。对最后一种情况，软件包将给出全部解。

在实际运用中，给定系统的方程组的系数一般是实数而非整数。我们假定，实系数中的小数部分的长度是有限的，并将

* 本文摘自《上海交通大学学报》，1986 年第 4 期。

其化为整数。这样，就可以借助于本软件包，用解整系数方程的方法来解给定的方程组。我们之所以将我们的工作限于解整系数方程，是基于如下的理由。

当给定系统的方程系数为整数而非实数时，软件包除了解方程之外，还可用于包括几何定理机器证明、几何关系的自动发现、多项式间关系之关系的判定等多种用途。我们将详细说明几个例子。

二、原理和方法

1. 多项式和多项式集合

除非特别指明，以下所提到的多项式由变量 x_1，……，x_n 和整系数组成。

一个多项式 A 中，若 x_c 在 A 中确实出现，但对任意的 $e>c$，x_e 不出现，则 A 在类 C 中，记为：cls（A）$=c$。若 A 非 0 而每个 x_i 都不出现，例如 A 是一个非 0 常数，则 A 在类 0 中，记为 cls（A）$=0$。

多项式 A 中 x_c 的次数 d 记为多项式 A 的次数：deg（A）$=d$。

在 cls（A）$=c>0$ 且 deg（A）$=d$ 的情况下，x_c^d 的系数本身是一个多项式 I，$I\neq 0$。这个多项式称为 A 的初式，记为 init（A）$=I$。

给定两个多项式 A 和 B，如果，cls（A）$<$cls（B），或者 cls（A）$=$cls（B）>0，但 deg（A）$<$deg（B），则称 A 的秩低于 B 的秩，或者称 B 的秩高于 A 的秩，记为 $A\leqslant B$ 或 $B\geqslant A$。

若 A 的秩不低于 B 的秩且 B 的秩亦不低于 A 的秩，则 A 和 B 为同秩，或者说两者等价，记为 $A\Leftrightarrow B$。

设 A：cls（A）$=c>0$ 且 deg（A）$=d$，若多项式 B 的 x_c 的次数（不是 deg（B））小于 d，则称 B 对 A 已约化。

设 A：cls（A）$=c>0$ 且 deg（A）$=d$，init（A）$=I$，对任意一个多项式 B，可决定唯一的、最小的整数 $s\geqslant 0$，使下面关系成立：

$$I^s * B = Q * A + R, \qquad \text{(REM)}$$

其中 Q 和 R 都是多项式，而 R 对 A 已约化。多项式 R 称为 B 对于 A 的余式，公式（REM）称为余式公式，整个求 R 的过程称为 B 对 A 约化的过程。

非 0 多项式序列 A_1，……，A_r，记为（ASC-A）。若条件（a）或（b）成立：

（a）$r=1$ 且 cls（A_1）$=0$；

（b）$0<$cls（A_1）$<$cls（A_2）$<\cdots<$cls（Ar）且对于 $\forall j>i$ 都对 A_j、A_i 约化。则称 ASC-A 是升列。

对（a）这种情况，（ASC-A）称为矛盾的，对（b）这种情况称普通的。

假定给出第二个升列（ASC-B）：B_1，…，B_s。若 ASC-A 与 ASC-B 适合条件（a）′或（b）′

（a）′$r\leqslant s$ 且 $\exists j\leqslant r$，使 $A_j\ll B_j$ 而 $\forall i<j$ 均有 $A_i\Leftrightarrow B_i$

（b）′$r>s$ 且对 $\forall i\leqslant s$ 均有 Ai⇔Bi

则称 ASC-A 的秩低于 ASC-B 的秩，或者说 ASC-B 的秩高于 ASC-A 的秩，记为（ASC-A）≪（ASC-B）或（ASC-B）≫（ASC-A）。

若（ASC-A）的秩不高于（ASC-B）的秩，而（ASC-B）的秩亦不高于（ASC-A）的秩，则称两者同秩或等价，记为：（ASC-A）⇔（ASC-B）。

若一个多项式 R 对任意一个普通升列 ASC-A 中的多项式 A_i 都已约化，则称多项式对 ASC-A 已约化。

对任一普通升列 ASC-A（如上）和任意的多项式 P，通过对序列 A_r，…，A_2，A_1 的约化，可得到下面的表达式

$$J * P = \mathrm{Sum}\ Q_i * A_i + R, \qquad \text{REM-A}$$

其中：J 是多项式A_i 的初始多项式l_i 的幂积，Q_i 和 R 都是多项式而 R 对 ASC-A 已约化，Sum 是对 $i=1$，2，…，r 求和的意思。多项式 R 称为 P 对于 ASC-A 的余式。公式 REM-A 是相应的余式公式，求 R 的过程称为 P 对升列 ASC-A 的约化。

一个有限的多项式集合称为 POL-SET（PS）。对于一个给定的 PS 集，可构造一些由 PS 集中多项式构成的最低秩升列。所有这些升列都是等价的，其中任意一个都称为 PS 的基本集。如果基本集是矛盾的，则 PS 本身也称为矛盾的。

2. 基本原理

下面仅叙述解方程系统的方法。对下述原理的证明可参阅[1]—[6]。

现在考虑引言中的方程集合 EQN-A，并用 PS-A 来记方程左边的多项式集合。本方法要从 PS-A 中推出一个升列 ASC-C，ASC-C 或者是矛盾的，或者是一个普通的：

$$C_1，C_2，\cdots\cdots，C_s。$$

在矛盾的情况中，给定的(PS)=(PS-A)，它的 Zero（PS-A）为空集，而在普通的情况中有

Zero（ASC-C/J） $-<$Zero（PS） $-<$Zero（ASC-C）

其中 J 是 ASC-C 中的 C_i 初式之积，$>-$表示包含，$-<$被包含。

不论是哪一种情形，升列 ASC-C 都称为 PS 的特征集，记为 Char（PS）。

在普通的情况下，用（Char）表示 Char（PS-A），对每个 i，$i=1$，2，…，S，将 $I_i=\mathrm{init}$（C_i）添加到（PS－A）中得

到的多项式集记为（PS_i），有

$$\text{Zero (PS)} = \text{Zero (Char/Ro)} + \text{Sum Zero } (PS_i), \qquad \text{(Zero}^*\text{)}$$

其中 Sum 是对 $i=1, 2, \cdots, S$ 求和，Ro 是 J 对（Char）的余式。现在我们轮流处理每个有限集（PS_i）如前，并除去所有的矛盾集合，这样就可在有限步内得如下形式的分解

$$\text{Zero (PS)} = \text{Sum Zero } (ASC_j / R_j)$$

其中每个（ASC_j）都是升列，R_j是一些对 ASC_j已约化的非零多项式。Sum 是对某些 j 求和。

在（PS-A）矛盾的情况下，方程系统（EQN-A）无解。否则，有解必被分解公式（Zero）描述：如果在（Zero）中出现的升列（ASC_j）的多项式数目少于（EQN-A）的变量数，则（EQN-A）有无限多个解。相反的情况是每个（ASC_j）包含正好 n 个多项式 A_{j_1}，…，A_{j_n}，且每个 A_{j_k}是由 $x_1 \cdots x_k$ 组成的多项式，但不包含任何 x_m（$m>k$），而 x_k 确实在 A_{j_k}中出现。我们可以用下述方法得到（EQN-A）的有限个解集。

用任何已知的方法解下列方程

$$A_{j_1}(x_1)=0,$$

$$A_{j_2}(x_1, x_2)=0,$$

$$\cdots\cdots\cdots\cdots\cdots\cdots$$

$$A_{j_n}(x_1, \cdots, x_n)=0,$$

依次解出 x_1，…，x_n，并从中选出 $R_j \neq 0$ 的解。如只需要实数解，可用我国古代的解法。

3．过程

我们现已知道如何解（EQN-A）方程了。解法基于 2 中的分解公式（Zero＊）。

解方程（EQN-A）的过程 S：

第 1 步：形成有限集（PS-A）=（PS）的特征集（char），（PS）与（EQN-A）相当。

若（Char）是矛盾的，则（EQN-A）无解，过程终止。

若（Char）是普通的，多项式的个数小于 n，则方程（EON-A）有无限多个解，过程终止。

否则进入下一步。

第 2 步：设（Char）为

$$C_1,\ C_2,\ \cdots,\ C_n,$$

其中 C_i 是由 x_1，…，x_i 组成的多项式，x_i 确实出现。解方程 $C_i=0$，$i=1$，2，…，n，依次解出 x_1，…，x_n。检查一下这些解是否使某个 C_i 的初始多项式 I_i 为 0，保留那些不使任何 I_i 为 0 的解。

第 3 步：仅考虑那些不为常数的 I_i，将 I_i 加入到有限集（PS）中，得到一新的有限集（PS_i），返回第一步，将 PS_i 当作 PS 进行处理。

整个过程必定在有限步内完成。

上述过程 S 中，最困难的是第一步中形成（PS）的特征集（Char）。这一步的做法如下：

过程 C 形成（Char）=Char（PS）：

第 1 步：形成有限集（PS）的基本集（BS）。

第 2 步：形成在（PS）中，但不在（BS）中的多项式的余式 R。

第 3 步：若有余式不为 0，则加入到（PS）中，将扩大后的（PS）仍称（PS）集，返回第一步。

第 4 步：若所有从第 2 步产生的余式均为 0，则基本集就是要找的特征集，过程终止。

过程 C 中最困难的是第 1 步，做法如下。

从（PS）形成基本集（BS）的过程 B：

第 1 步：从有限集（PS）中找一个秩最低的多项式 B_1。

第 2 步：若 B_1 是常数多项式，则 B_1 单独形成基本集，过程终止，有限集（PS）是矛盾的。

第 3 步：设 B_1，…，B_K是一个已从（PS）中选出的升列（ASC），然后测试一下在（PS）中是否还有对（ASC）已约化而不在（ASC）中的多项式存在，若没有，则（ASC）就是所需的基本集（BS），过程终止。否则，进入下一步。

第 4 步：从（PS）中选出不在（ASC）中且对（ASC）已约化而秩最低的多项式，将该多项式加入（ASC），并重新命名所得升列为（ASC）。返回第 3 步。

整个过程必在有限步内终止。

4. 修改

一般而言，给定的有限集（PS）中仅少数几个多项式，且每个多项式的项数却很少，但在过程中，多项式的项数和多项式的个数增加极快，使普通的计算机不能胜任所需的运算。为了避免这个问题，我们对某些概念作如下修改：

（1）在升列的定义中，条件（b）可降为：A_j 仅对那些A_i（$i<j$）已约化，其中 x_{c_i}（$C_i=\mathrm{cls}(A_i)$）在 A_j 的初式中出现。我们来区别一下这两个概念，前者是升列的强定义，后者是弱定义。对于以升列概念为基础的基本集和特征集概念，也采用同样的定义。

（2）不论是强升列定义或弱升列定义，有限集（PS）的特征集（Char）基本上都是遵循下面的模式求得的

$$
\begin{aligned}
&(\mathrm{PS})=(\mathrm{PS}_1)\rightarrow(\mathrm{PS}_2)\rightarrow\cdots\cdots\rightarrow(\mathrm{PS}_m)\\
&(\mathrm{BS})=(\mathrm{BS}_1)\rightarrow(\mathrm{BS}_2)\rightarrow\cdots\cdots\rightarrow(\mathrm{BS}_m) \qquad (\mathrm{SCH})\\
&\qquad\quad(\mathrm{RS}_1)\rightarrow(\mathrm{RS}_2)\rightarrow\cdots\cdots\rightarrow(\mathrm{RS}_m)
\end{aligned}
$$

在模式中每个（BS_i）是有限集（PS_i）的基本集，每个（RS_i）是（PS_i）的但不在（BS_i）中的，对 BS_i 的非零余式

集；每个（PS_i）是有限集（PS_i）与（RS_j），$j=i-1$ 的并集。换言之，

$$(PS_i)=(PS_j)+(RS_j), j=i-1。$$

整个过程在（RS_m）= 空集时终止，（BS_m）是所需的特征集（Char）。

在这个模式中，有限集（PS_i）将逐步增大，（RS_i）的形式也很复杂。为了避免这个困难，我们对模式进行如下修改。

我们仅将每个（PS_i）的求法改变，即

$$(PS_i)=(BS_j)+(RS_j), j=i-1,$$

过程仍在（RS_m）为空集时终止，相应的 BS_m 称为半特征集。将（BS_m）加入原来的有限集（PS），得到新的有限集，仍记为（PS），即

$$(PS)=(PS)+(BS_m)。$$

然后重复同一过程如前，直至我们得到 $m=1$，这时过程终止，最后的基本集（BS）是所需的特征集。

我们还可以在过程 S 的第一步中用半特征集替代特征集，从而进一步简化过程 S。我们的程序设计是基本简化和修改后的过程。

三、程序设计

1. 总体描述

共 4 个程序，表格中给出了各程序的说明。

程　序　名	输　　入	输　　出	作　　用
Solv-pmat.for	字符串形式的有限集，放在文件#3＝'polyms'中	有限集以矩阵形式存放在文件：#4＝'matrices'中	将有限集从字符串形式转换成矩阵形式
Solv-mpol.for	矩阵形式的有限集，放在文件：#3＝'matrices'	字符串形式的有限集，放在文件#4＝'polyms'中	将有限集从矩阵形式转换成字符串形式
Solv-Chrst.for	矩阵形式的有限集放在文件：#3＝'polset'中	特征集在文件：#4＝'Char-set' 因子集在文件：#5＝'init-fctr-set' 初式集亦在文件#5 过程在文件：#6＝'Procedure'	从(PS)中确定：特征集Char，初集init以及因子集fctr
Solv-eqns.for	矩阵形式的升列在文件#3＝'asc-set'	给出一组方程的复杂的解	给出阶梯形式方程组的解

在上表中，文件'init-fctr-set'中包含特征集的初式，前者在文件'Char-set'中。另外还包含一个因子集 Factor-set，

此因子集系在解的过程中，为了避免过于复杂而消去的系数公因子 x_K 组成。因有可能，在 x_K 为 0 的情况下，丢失一些解，故必须在下述过程中重新检验。

用这些程序解一个给定方程系统（EQN-A）的过程如下：

首先从一个给定的方程系统（EQN-A）中形成有限集（PS-A），将（PS-A）放入文件 ‘polyms’，作为程序 Solv-pmat. for 的输入，输出在文件 ‘matrices’ 中，再放入文件 ‘polset’ 中，作为程序 Solv-Chrst. for 的输入。若输出的文件 ‘Char-set’ 中仅一个常数多项式或者多项式的个数少于变量的个数，则我们将不再考虑而过程停止。否则，将结果放入文件 ‘asc-set’ 作为程序 Solv-eqns. for 的输入。输出是（EQN-A）的有限个解。若有非常数的初式，例如 I_K，在文件 ‘init-fctr-set’ 中出现，我们将每个 I_K 添加到（PS-A），从而得到一个新的有限集，仍记作（PS-A）。我们将重新进行上述过程，进一步求得原来（EQN-A）更多的解。如果在 ‘init-fctr-set’ 有某一 x_K，则添加入（PS-A），命所得新的有限集仍记为（PS-A）而如前进行。这时我们就求出了 $x_K=0$ 的（EQN-A）的所有的可能的解。这样，我们就得到了结果，（EQN-A）是否无解、无穷多解、或者有限个解。对最后一种情况，给出 EQN-A 的全部解。

2. 多项表达式

多项式可用字符串形式或矩阵形式表示。

矩阵形式的多项式可用 4 个整数数组表示：

iind, ivar, iexp, icof。

iind = 5 个整数（it, iv, ic, id, ip）的索引集，

it = 多项式中实际项数，

iv = 多项式中的实际变量数，

ic = 类别，即多项式中 x_c 实际出现的最大 c 值，

id = 多项式的 x_c 的阶，

ip = 多项式表示数组 ivar 中 x_c 出现的位置。

ivar = 多项式中实际出现的变量 xi 的集合，次序是随意的，但一经确定就不变。

iexp = 多项式相继各项的 ivar 的变量的指标集合。

icof = 多项式相继各项的系数集合。

3．变量、数组和控制参数

在程序中用了一些全局性的标记。

（1）变量

np = 各有限集多项式的虚拟个数，

nt = 各有限集的多项式各项的个数，

nv = 各有限集的变量的各项的个数，

kp = 各有限集多项式的实际个数，

kv = 各有限集变量的实际个数。

（2）数组

jind（np，5）= np 多项式的 iind 集合，

jvar（np，nv）= np 多项式的 ivar 集合，

jexp = np 多项式的 iexp 集合，

jeof = np 多项式的 icof 集合。

iexp 数组是 2 维数组，nt 行 nv 列，这是在 Solv-pmat．for 程序中的情况，但在 Solv-chrst．for 程序中，用的是 1 维数组，以节约空间。这时，iexp 中对应多项式第 i 项，ivar 中第 j 个变量位置是（（$i-1$）* kv + j）。jexp 与 jcof 数组同理。

在程序 Solv-chrst．for 中可有两个限集（PS-A）和（PS-B）。相应的数组分别用 iaind 与 ibind 等进行区分。

（3）控制参数

有一些控制参数从终端读入，它们由用户决定。

iwt = 0～显示于 CRT,

iwt = k≠0～写入文件$_{\#}$K,

iqt<0～多项式不打印,

iqt>0～多项式以数组形式打印,

ips = 0～在指定位置停止,

ips≠0～在指定位置继续,

ired = 1～集合以弱意义出现,

ired = 2～集合以强意义出现,

init≠0～要形成初式,

init = 0～不要形成初式,

ifct≠0～除去单变量公因子,

ifct = 0～不除去单变量公因子。

4．程序 Solv-pmat．for & Solv-mpol．for

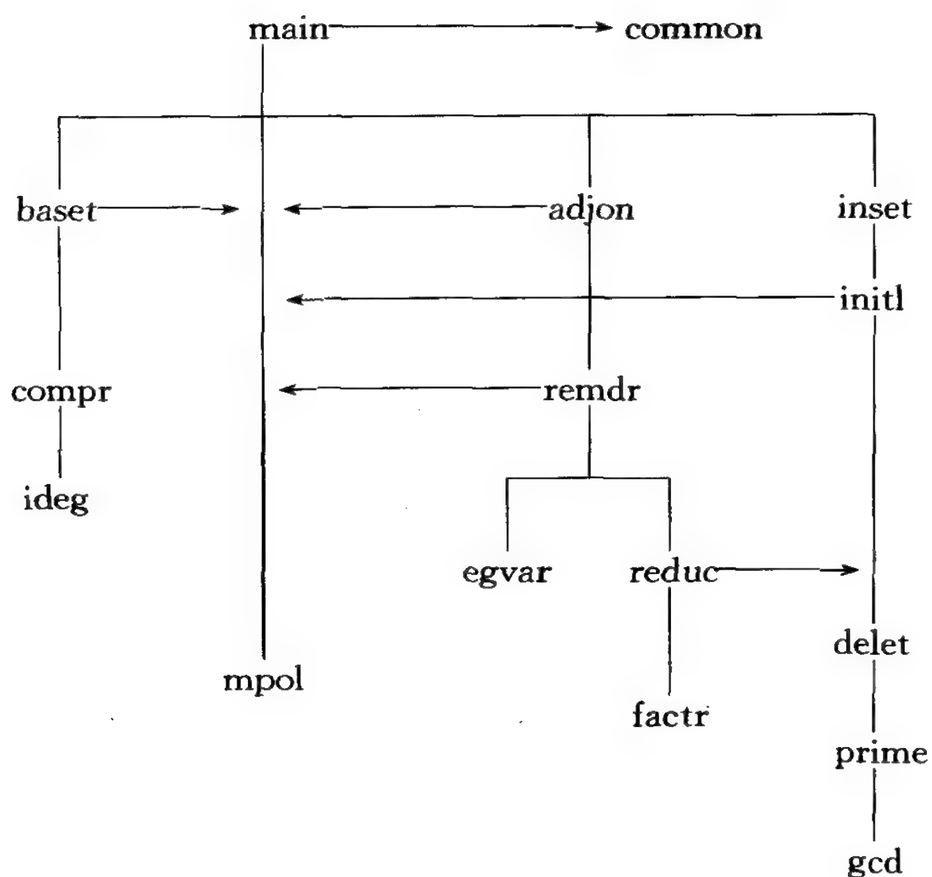
这两个程序相对地简单，可在需要时用于任何目的。

5．程序 Solv-chrst．for

这是整个软件包的主程序。

先从文件$_{\#}$3 = ‘polset’中读入有限集(PS-A),然后进行 B 中所述的修改和简化过程,最后的结果是存放在文件$_{\#}$4 = ‘Char-set’中的半特征集、存放在文件$_{\#}$5 = ‘init-fctr-set’中的初始化多项式集因子集。若 iwt = 0 这个参数输入,则整个执行过程存入文件$_{\#}$6 = ‘procedure’。

(1) 子程序结构。



（2）主要子程序作用

子程序	作　　　　　　用
baset	从有限集 pol-set(PS-A)中选择基本集(BS-B)。
adjon	若有非 0 余式，则加入(BS-B)，形成新的有限集(PS-A)。
remdr	将多项式 A、B 中的 iavar 和 ibvar 混合成相同的变量数组并将 A、B 作相应的改变。
reduc	将多项式 A 对多项式 B 约化。
inset	形成相应于基本集的初式集。

6. 程序 Solv-eqns. for

可利用各种已有程序，从略。

参 考 文 献

[1] J. F. Ritt, Differential equations from the algebraic standpoint, AMS 1932.

[2] J. F. Ritt, Differential algebra, AMS 1950.

[3] Wu Wen-tsun, On the decision problem and the mechanization of theoremproving in elementary geometry, Scientia Sinica 21 (1978) 159 — 172, or in Contemporary mathematics vol. 29 (1984) 213-234.

[4] Wu Wen-tsun, Basic Principles of mechanical theorem proving in elementary geometries J. Sys. Sci. & Math. Sci. 4 (1984) 207-235.

[5] Wu Wen-tsun, Basic Principles of mechanical Proving of geometrical theorems, Part I. Elementary geometries, Science Press (1984).

[6] Wu Wen-tsun, On systems of algebraic equations—an application of Ritt Principle, Kexue Tong Bao (to be Published), 31 (1986) 1-5.

附　　录

On the decision problem and the mechanization of theorem-proving in elementary geometry*

Abstract

The idea of proving theorems mechanically may be dated back to Leibniz in the 17th century and has been formulated in precise mathematical forms in this century through the school of Hilbert as well as his followers on mathematical logic. The problem consists in essence in replacing qualitative difficulties inherited in usual mathematical proofs by quantitative complexities of calculations on standardizing the proof procedures in an algorithmic manner. Such quantitative complexi-ties of calculations, formerly far beyond the reach of human abilities, have become more and more trivial owing to the occurrence and rapid development of computers. In spite of vigorous efforts, however, researches in this direction give rise quite often to negative results in the form of un-

* Scientia Sinica 21 (1978) 159—172, re—published in Automated Theorem proving: after 25 Years, (Eds. W. W. Bledsoe & D. W. Loveland (1984) 235—242.

decidable mathematical theories. To cite a notable positive result, we may mention Tarski's method of proving theorems mechanically in elementary geometry and elementary algebra. The methods of Tarski as well as later ones are largely based on a generalization of Sturm theorem and are still too complicated to be feasible, even with the use of computers . The present paper, restricted to theorems with betweenness out of consideration and based on an entirely different principle, aims at giving a mechanical procedure which permits to prove quite non-trivial theorems in elementary geometry even by hands.

I. Formulation of the problem

A. Tarski in a classic paper [14] of 1948 has settled the decision problem of real closed field with one of its main aims to give mechanical proofs of theorems in elementary geometry. Alternative proofs of Tarski's result have later been given by Seidenberg, A. Robinson and P. J. Cohen, cf [12, 9, 2]. These authors have even suggested construction of certain decision machines to carry out such mechanical proofs. However, such a procedure seems to be far from being realized. In fact, only proofs of very trivial theorems in elementary geometry have actually been carried out on computers, cf. e. g. [6, 7]. The purpose of the present paper is , leaving aside questions involving betweenness of points, to give an alternative solution of the decision problem of elementary geometry based on a principle entirely different from those employed by the authors above-mentioned. Our method permits to furnish mechanical proofs of quite difficult geometrical theorems which can be practiced even by hands, i. e., by

means of papers and pencils only. The programming on a computer, based on such a method, though has not yet been done, will present no actual difficulties at all.

We shall restrict our considerations wholly to plane elementary geometry, though our method may be applied to the consideration of various other kinds of geometry. The first step of our method consists in the algebraic formalization of the geometrical problems involved. Points in the plane are to be defined as ordered pairs of numbers in a fixed field, say the field of rational numbers R. A dictionary is then set up turning geometrical relations into algebraic expressions which may be considered as either definitions or axioms. For example, for points $A_i = (x_i, y_i)$, distinct or not, we shall say:

A_1A_2 is parallel to A_3A_4

if $(x_1 - x_2)(y_3 - y_4) - (x_3 - x_4)(y_1 - y_2) = 0$,

A_1A_2 is orthogonal to A_3A_4

if $(x_1 - x_2)(x_3 - x_4) + (y_1 - y_2)(y_3 - y_4) = 0$,

the length-square of A_1A_2 is $r^2 = (x_1 - x_2)^2 + (y_1 - y_2)^2$, etc.

We may replace the basic field R by other fields, make correspondence the points to other kinds of number-sets, or modify the algebraic expressions in the axioms, e. g., instead of the length-square function r^2 given above, we consider the function $r^4 = (x_1 - x_2)^4 + (y_1 - y_2)^4$. We then go to other realms of geometry, non-euclidean geometry, real or complex projective geometry, finite geometry, etc. We shall however stick ourselves in what follows to plane elementary geometry only which

has some representative character.

To illustrate our method of treatment, let us cite first a simple example. Consider the following statement:

(S_n) Let $A_0A_1A_2$ be a right-angled triangle with right angle at A_0. If x_1, x_2 denote the lengths of sides A_0A_1, A_0A_2 and x_3 is the length of the hypotenuse, then

$$x_1^n + x_2^n = x_3^n.$$

The problem is to decide whether the statement (S_n) is true or not and to give an algorithmic procedure of proving or disproving (S_n) which holds good for all statements alike in elementary geometry (with betweenness out of consideration).

To solve this problem, let us remark first that the points, etc., occurring in the statement have a generic character subjected to the conditions implied in the hypothesis of the statement. Thus, if we represent the points in question in coordinates with $A_0 = (x_0, v_0)$, $A_1 = (u_1, v_1)$, $A_2 = (u_2, v_2)$, the coordinates v_0, u_1, v_1, u_2, v_2 can be considered as indeterminates. On the other hand, the other coordinates and geometric entities x_0, x_1, x_2, x_3 are then algebraically dependent on these indeterminates, being restricted by following algebraic equations according to hypothesis of the statement (S_n):

$$f_0 \equiv (u_1 - x_0)(u_2 - x_0) + (v_1 - v_0)(v_2 - v_0) = 0,$$
$$f_1 \equiv x_1^2 - (u_1 - x_0)^2 - (v_1 - v_0)^2 = 0,$$
$$f_2 \equiv x_2^2 - (u_2 - x_0)^2 - (v_2 - v_0)^2 = 0,$$
$$f_3 \equiv x_3^2 - (u_1 - u_2)^2 - (v_1 - v_2)^2 = 0.$$

The conclusion in the statement (S_n) is equivalent to

$$g_n \equiv x_3^n - x_1^n - x_2^n = 0.$$

Let us now take once and for all the rational number field R as the base field. Let A^9 be the affine space on R with coordinates $(v_0, u_1, v_1, u_2, v_2, x_0, x_1, x_2, x_3)$ arranged in that definite order. Then the above equations $f_i = 0$ define an algebraic variety V of dimension 5, in the present case irreducible over, R, with some generic point $(v_0, u_1, v_1, u_2, v_2, \bar{x}_0, \bar{x}_1, \bar{x}_2, \bar{x}_3)$ of which v_0, u_1, v_1, u_2, v_2 are indeterminates. The truth or untruth of the statement (S_n) amounts to $g_n \equiv 0$ or $g_n \not\equiv 0$ on V respectively.

It turns out that the general decision problem can be formulated in the following manner.

Problem. In a certain affine spaceA^n of dimension $n = r + d$ over R with coordinates $(u_1, \cdots, u_d, x_1, \cdots, x_r)$, consider an alge-braic variety V with defining equations (u_i being independent indeterminates)

$$(\mathrm{I}) \qquad \begin{cases} f_1 \ (u_1, \cdots, u_d, x_1) = 0, \\ f_2 \ (u_1, \cdots, u_d, x_1, x_2) = 0, \\ \cdots \\ f_r \ (u_1, \cdots, u_d, x_1, x_2, \cdots, x_r) = 0. \end{cases}$$

The variety V may eventually split into irreducible components, all of (real) dimension $\leqslant d$. Those of dimension $= d$, with generic points of the form $(u_1, \cdots, u_d, \bar{x}_1, \cdots, \bar{x}_r)$ for which $\bar{x}_j$ are alge-braic over the field $K = R \ (u_1, \cdots, u_d)$ will have a union V^* usually coincident with V. Let a polynomial g $(u_1, \cdots, u_d, x_1, \cdots, x_r)$ (or a set of such polynomials g_k) in $R\ [u_1, \cdots, u_d, x_1, \cdots, x_r]$ be given. It is to decide in an algorithmic manner whether

$$g \equiv 0$$

(or all $g_k \equiv 0$) on V^* or not.

In the above formulation the algebraic variety V, or preferably V^* reflects the hypothesis of the geometric statement considered. Either V or V^* will be called the *associated variety* of the state-ment in question. The variety V^* considered as one defined on the field$K = R\ (u_1, \cdots, u_d)$ is of dimension 0. The form of equations (I) shows that the algebraically dependent variables$x_1, \cdots, x_r$ are to be adjoined to K *successively* which reflects the geometrical fact that, starting from some generic points on certain generic lines, circles, etc., new points are to be *successively* adjoined in an algebraic manner by various geometric operations of joining points, drawing parallels, perpendiculars or circles, forming inter-sections of lines and circles, etc. These geometrical constructions give rise to algebraic equations involving x and u which can easily be turned into the form (I) be simple elimination procedure. In fact, the starting equations in x and u are rarely higher than 2. We shall call the variables u as the *parameters* and x as the *dependents*. It is also to be remarked that the condition for all components of V to be of dimension $\leqslant d$ over R reflects just the *determinate* character of the geometric statements to be considered, and the restriction to V^* reflects the depriving of degen-eracies in our consideration. Both of these are, in reality, implicitly implied in the hypothesis of ordinary geometrical theorems. On the other hand, the equation $g \equiv 0$ on V^* (or set of $g_k \equiv 0$ on V^*) is the algebraic equivalent of the conclusion of the statement to be proved or disproved. We shall call g or set of g_k in what follows the *deciding polynomial* (*s*). of the geometrical statement in question.

Theoretically, the methods given by Hermann in [4] permit already to solve the above decision problem in an algorithmic manner. However, his methods are so complicated to give rise to astronomical expansions that even the simplest geometrical theorems can hardly be proved . On the contrary the decision procedure given below takes advantage of the particular character of the equations (I) and permits to prove mechanically quite nontrivial theorems even by hands, i. e. by means of pencil and paper only.

Our method of decision procedure is based on the following three theorems:

Theorem 1 *There is an algorithmic procedure permitting us to split the associated variety V^* of any determinative geometric statement defined by (I) into subvarieties V' irreducible over R each of which, considered as defined over the field* $K = R(u_1, \cdots, u_d)$ *has a representative basis (i. e. basis of the associated prime ideal) of. the form*

$$(p_1, \cdots, p_r),$$

having the following properties:

$(T1)_1$. *Each p_i is a polynomial in $R[u_1, \cdots, u_d, x_1, \cdots, x_i]$ of some degree $m_i > 0$ in x_i*

$(T1)_2$. *The coefficients of p_i considered as a polynomial in x_i, are polynomials in $R[u_1, \cdots, u_d, x_1, \cdots, x_{i-1}]$ having no common factor and with degree in x_j less than m_j for $j = 1, \cdots, i-1$.*

$(T1)_3$. *The leading coefficient of p_i, considered as a polyno-mial in x_i, is a polynomial $\neq 0$ in $R[u_1, \cdots, u_d]$ free of all x.*

$(T1)_4$. *p_1, as a polynomial in x_1, is irreducible in the*

field $K = R\ (u_1, \cdots, u_d)$, *and for each* $i > 1$, p_i *as a polynomial in* x_i, *is irreducible in the field obtained by adjoining* $x_1, \cdots, x_{i-1}$ *to K by the algebraic equations* $p_1 = 0, \cdots, p_{i-1} = 0$.

It is clear that the polynomials $p_1, \cdots, p_r$ are uniquely determined by V' up to multipliers in R and will be said to form a *privileged basis* of V' (more exactly, of the prime ideal associated to V' over K), with respect to the given order $x_1, \cdots, x_r$ of the dependents. Remark that the notion is in reality due to Gröbner under the name of prime basis, cf. e. g. [3] and cf. also [8] for the intimately related concept of characteristic sets introduced by R. F. Ritt.

Theorem 2 *Let* $(p_1, \cdots, p_r)$ *be a privileged basis of any irreducible component* V' *of the associated variety* V^*. *There is an algorithmic procedure which permits us to determine, for any polynomial h in* $R\ [u_1, \cdots, u_d, x_1, \cdots, x_r]$, *an equation of the form*

$$Dh = \sum_{i_3=0}^{m_3-1} h_{i_1}, \cdots, i_r x^{i_1} \cdots x^{i_r} + \sum_{i=1}^{r} A_i p_i,$$

verifying the following conditions:

$(T2)_1$. D, $h_{i_1} \cdots i_r$ *are all polynomials in* $R\ [u_1, \cdots, u_d]$ *and* $D \neq 0$,

$(T2)_2$. A_i *are all polynomials in* $R\ [u_1, \cdots, u_d, x_1, \cdots, x_r]$.

The polynomials $h_{i_1}, \cdots, i_r$ which are uniquely determined up to multipliers in R by the algorithmic procedure, will be called the *remainder constituents* of the polynomial h with respect to the privileged basis $(p_1, \cdots, p_r)$ of V', or by abuse of language, simply

the remainder constituents of h with respect to V'

Theorem 3. *For a geometrical statement with associated variety V^* and deciding polynomial g (or a set g_k of deciding polynomials) to be true, it is necessary and sufficient that for any irreducible component V' of V^*, all remainder constituents of g (or of all g_k) should be identically zero.*

Ⅱ. Examples

Before giving proofs of these theorems in IV, we shall illustrate their use by some examples below.

Ex. 1. For the geometrical statement (S_n) about right-angled triangles as cited in the beginning of the present paper, we see readily that the associated variety V in parameters v_0, u_1, v_1, u_2, v_2 and dependents x_0, x_1, x_2, x_3, is already irreducible over R and possesses a privileged basis (p_1, p_2, p_3, p_4), where

$$p_1 = x_0^2 - (u_1 + u_2)x_0 + u_1u_2 + (v_1 - v_0)(v_2 - v_0),$$
$$p_2 = x_1^2 + (u_1 - u_2)x_0 - u_1^2 + u^1u_2 - (v_1 - v_0)(v_1 - v_2),$$
$$p_3 = x_2^2 - (u_1 - u_2)x_0 - u_2^2 + u_1u_2 + (v_2 - v_0)(v_1 - v_2),$$
$$p_4 = x_3^2 - (u_1 - u_2)^2 - (v_1 - v_2)^2.$$

It is readily verified that

$$g_2 = p_4 - p_2 - p_3,$$

while for $n > 2$, we have

$$g_n \not\equiv 0 \bmod (p_1, p_2, p_3, p_4).$$

It follows that (S_n) is true only for $n = 2$ which corresponds to

the Keu-Kou Theorem.

Ex. 2. Our decision procedure can also be applied to give mechanical proofs of trigonometric identities. Consider, e.g. the following statements:

(S) If $A_1 + A_2 + A_3 = 180°$, then

$$\sin 2A_1 + \sin 2A_2 + \sin 2A_3 = 4\sin A_1 \sin A_2 \sin A_3.$$

(C) If $A_1 + A_2 + A_3 = 180°$, then

$$\cos 2A_1 + \cos 2A_2 + \cos 2A_3 + 4\cos A_1 \cos A_2 \cos A_3 = 0.$$

To decide whether (S) or (C) is true, let us set

$$\sin A_i = s_i, \quad \cos A_i = c_i, \quad (i = 1, 2),$$

$$\sin 2A_i = x_i, \quad \cos 2A_i = y_i, \quad (i = 1, 2, 3),$$

$$\sin A_3 = z_1, \quad \cos A_3 = z_2.$$

Take c_1 and c_2 as parameters and

$$s_1, \ s_2, \ z_1, \ z_2, \ x_1, \ x_2, \ x_3, \ y_1, \ y_2, \ y_3$$

as dependents (in this order), then the associated variety of (S) or (C) is already irreducible in R and possesses a privileged basis $(p_1, \cdots, p_{10})$ given by

$$p_1 = s_1^2 + c_1^2 - 1,$$

$$p_2 = s_2^2 + c_2^2 - 1,$$

$$p_3 = z_1 - c_2 s_1 - c_1 s_2.$$

$$p_4 = z_2 - s_1 s_2 + c_1 c_2,$$

$$p_5 = x_1 - 2c_1 s_1,$$

$$p_6 = x_2 - 2c_2 s_2,$$

$$p_7 = x_3 - 2z_1 z_2,$$

$$p_8 = y_1 - 2c_1^2 + 1,$$

$$p_9 = y_2 - 2c_2^2 + 1,$$

$$p_{10} = y_3 + 4c_1 c_2 s_1 s_2 - 4c_1^2 c_2^2 + 2c_1^2 + 2c_2^2 - 1.$$

The deciding polynomials of the statements (S) and (C) are given respectively by

$$g_s = x_1 + x_2 + x_3 - 4s_1s_2z_1,$$
$$g_c = y_1 + y_2 + y_3 + 4c_1c_2z_2.$$

We verify readily that

$$\begin{aligned} g_s &= -2c_2s_2p_1 - 2c_1s_1p_2 - 2p_3\ (s_1s_2 + c_1c_2) \\ &\quad + 2z_1p_4 + p_5 + p_6 + p_7 \\ &\equiv 0 \bmod (p_1, \cdots, p_{10}), \\ g_c &= -1 + 4c_1c_2p_4 + p_8 + p_9 + p_{10} \\ &\not\equiv 0 \bmod (p_1, \cdots, p_{10}. \end{aligned}$$

Hence (S) gives a true identity while (C) does not.

Ex. 3 Let us consider the Simson-Line Theorem which corresponds to the following statement:

(S_s). From a point A_4 on the circumscribed circle of a triangle $A_1A_2A_3$, perpendiculars are drawn to the sides of the triangle. Then the feet of the perpendiculars are in a line.

To prove this, let us take for simplicity the center of the circumscribed circle as the point $(0, 0)$, while the radius is r. Let the points A_i $(i = 1, 2, 3, 4)$ be (x_i, u_i) and the feet of perpendiculars A_i be (x_i, y_j), $j = 5, 6, 7$. Consider r, u_1, u_2, u_3, u_4 as parameters and

$$x_1,\ x_2,\ x_3,\ x_4,\ y_5,\ y_6,\ y_7,\ x_5,\ x_6,\ x_7$$

as the dependents in the order indicated.

A prime basis $(p_1, \cdots, p_{10})$ of the associated Simson variety which is irreducible is readily given as follows:

$$p_1 = x_1^2 + u_1^2 - r^2,$$
$$p_2 = x_2^2 + u_2^2 - r^2,$$
$$p_3 = x_3^2 + u_3^2 - r^2,$$
$$p_4 = x_4^2 + u_4^2 - r^2,$$
$$p_5 = 2r^2 y_5 - h_5,$$
$$p_6 = 2r^2 y_6 - h_6,$$
$$p_7 = 2r^2 y_7 - h_7,$$
$$p_8 = u_{23} x_5 - x_{23} y_5 + u_3 x_2 - u_2 x_3,$$
$$p_9 = u_{31} x_6 - x_{31} y_6 + u_1 x_3 - u_3 x_1,$$
$$p_{10} = u_{12} x_7 - x_{12} y_7 + u_2 x_1 - u_1 x_2.$$

In the above equations we have set for simplicity:

$$u_{ij} = u_i - u_j, \quad x_{ij} = x_i - x_j, \quad (i, \ j = 1, \ 2, \ 3),$$
$$h_5 = u_4 x_2 x_3 - (u_2 x_3 + u_3 x_2) x_4 + r^2 (u_2 + u_3 + u_4) - u_2 u_3 u_4.$$

Similarly for h_6 and h_7.

The deciding polynomial is given by

$$g = x_5(y_6 - y_7) + x_6(y_7 - y_5) + x_7(y_5 - y_6).$$

Straightforward calculation shows again

$$g \equiv 0 \quad \mathrm{mod} \ (p_1, \ \cdots, \ p_{10}),$$

which proves the truth of Simson statement (S_s).

Ex. 4. For a less trivial example, let us consider the Feuerbach theorem which corresponds to the following statement:

(S_F) The 9-point circle of a triangle is tangent to the four inscribed and the escribed circles of the triangle.

Let us take the three vertices of the triangle as $(2u_i, \ 2v_i)$, the center and radius of the 9-point circle as $(x_1, \ y_1)$ and r_1,

and the center and radius of either the inscribed or any of the escribed circles as (x_2, y_2) and r_2. Introduce also variables z_1, z_2, z_3 corresponding to lengths of the sides of the triangle. Then with $u_1, v_1, u_2, v_2, u_3, v_3$ as the parameters and $z_1, z_2, z_3, x_1, y_1, x_2, y_2, r_1, r_2$ as dependents in this order, the associated Feuerbach variety V_F splits into irreducible ones having privileged basis $(p_1, \cdots, p_9)$ given by:

$$p_1 = z_1^2 - u_{23}^2 - v_{23}^2,$$
$$p_2 = z_2^2 - u_{31}^2 - v_{31}^2,$$
$$p_3 = z_3^2 - u_{12}^2 - v_{12}^2,$$
$$p_4 = 4\Delta x_1 - \alpha_1,$$
$$p_5 = 4\Delta y_1 - \beta_1,$$
$$p_6 = 2\Delta x_2 - \alpha_2,$$
$$p_7 = 2\Delta y_2 - \beta_2,$$
$$p_8 = 4\Delta r_1 - z_1 z_2 z_3,$$
$$p_9 = 2\Delta r_2 + \gamma.$$

In the above formulas we have put for simplicity

$$u_{ij} = u_i - u_j, \quad v_{ij} = v_i - v_j, \quad (i, j = 1, 2, 3),$$

$$2\Delta = \begin{vmatrix} u_1, & v_1, & 1 \\ u_2, & v_2, & 1 \\ u_3, & v_3, & 1 \end{vmatrix},$$

$$\alpha_1 = v_{23} v_{31} v_{12} + u_1^2 v_{23} + u_2^2 v_{31} + u_3^2 v_{12} - 2(u_2 u_3 v_{23} + u_3 u_1 v_{31} + u_1 u_2 v_{12}),$$

$$\beta_1 = -u_{23} u_{31} u_{12} - v_1^2 u_{23} - v_2^2 u_{31} - v_3^2 u_{12} + 2(v_2 v_3 u_{23} + v_3 v_1 u_{31} + v_1 v_2 u_{12}),$$

$$\alpha_2 = -[v_{23} v_{31} v_{12} + v_1(u_2^2 - u_3^2) + v_2(u_3^2 - u_1^2) + v_3(u_1^2 -$$

$$u_2^2) + \varepsilon_2\varepsilon_3 v_{23} z_1 z_3 + \varepsilon_3\varepsilon_1 v_{31} z_3 z_1 + \varepsilon_1\varepsilon_2 v_{12} z_1 z_2\},$$

$$\beta_2 = u_{23} u_{31} u_{12} + u_1(v_2^2 - v_3^2) + u_2(v_3^2 - v_1^2) + u_3(v_1^2 - v_2^2)$$
$$+ \varepsilon_2\varepsilon_3 u_{23} z_2 z_3 + \varepsilon_3\varepsilon_1 u_{31} z_3 z_1 + \varepsilon_1\varepsilon_2 u_{12} z_1 z_2,$$

$$\gamma = \varepsilon_1(u_{31} u_{12} + v_{31} v_{12}) z_1 + \varepsilon_2(u_{12} u_{23} + v_{12} v_{23}) z_2$$
$$+ \varepsilon_3(u_{23} u_{31} + v_{23} v_{31}) z_3 + \varepsilon_1\varepsilon_2\varepsilon_3 z_1 z_2 z_3.$$

Remark that the choice of $\varepsilon_i = +1$ or -1 corresponds to the 4 inscribed or escribed circles and reflects the reducibility of the Feuerbach variety.

Now the deciding polynomial of statement (S_F) is of the form

$$g_\eta = (\eta r_1 - r_2)^2 - (x_1 - x_2)^2 - (y_1 - y_2)^2,$$

in which $\eta = +1$ or -1. By a straightforward calculation, however lengthy and tedious, we verify that g_η will be a linear combination of $p_1, \cdots, p_9$ for $\eta = +\varepsilon_1\varepsilon_2\varepsilon_3$ but not so for $\eta = -\varepsilon_1\varepsilon_2\varepsilon_3$. Thus the Feuerbach statement is a true theorem and $\eta = +\varepsilon_1\varepsilon_2\varepsilon_3$ reflects the manner of contact between the 9-point circle and the respective in-or es-cribed circle. Remark that in our formulation we have left aside the question of betweenness and the manner of contact of circles is not of interest to us. In this way we can take $g = g_{+1} \cdot g_{-1}$ as the deciding polynomial if we like. We may also take the Feuerbach variety in the affine space of u_i, v_i, x_j, y_j Only without the introduction of z_1, z_2, z_3 so that it is irreducible at the outset.

Ⅲ. Some lemmas

To make some preparations we shall consider a field $R\ (u_1, \cdots, u_d, x_1, \cdots, x_r)$ in which $u_1, \cdots, u_d$ are transcendental

while $x_1, \cdots, x_r$ are algebraic over the base field R. The algebraic extensions to $x_1, \cdots, x_r$ are defined successively by the following equations

$$
(\mathrm{II}) \begin{cases} p_1(x_1) \equiv p_{10} x_1^{m_1} + p_{11} x_1^{m_1 - 1} + \cdots + p_{1m_1} = 0, \\ p_2(x_1, x_2) \equiv p_{20} x_2^{m_2} + p_{21} x_2^{m_2 - 1} + \cdots + p_{2m_2} = 0, \\ \cdots\cdots \\ p_r(x_1, \cdots, x_r) \equiv p_{r0} x_r^{m_r} + p_{r1} x_r^{m_r - 1} + \cdots + p_{rm_r} = 0. \end{cases}
$$

It is assumed that for $1 \leqslant i \leqslant r$, the p_{ij}'s are polynomials in the ring $p_{i-1} = R\ [u_1, \cdots, u_d, x_1, \cdots, x_{i-1}]$, that p_{i0} are polynomials $\neq 0$ in $p_0 = R\ [u_1, \cdots, u_d]$, and that p_i considered as polynomials in x_i are irreducible over the field $K_{i-1} = R\ (u_1, \cdots, u_d, x_1, \cdots, x_{i-1})$ defined by the equations $p_1(x_1) = 0, \cdots, p_{i-1}(x_1, \cdots, x_{i-1}) = 0$. We put here

$$K_0 = K = R\ (u_1, \cdots, u_d).$$

let r be the collection of sets of indices $I = (i_1, \cdots i_r)$ with $0 \leq i_j \leq m_j - 1$. For such an I we shall write symbolically

$$x^I = x_1^{i_1} \cdots x_r^{i_r}.$$

Any polynomial of the form

with coefficients a_I in a certain ring of field F will be called then a *normalized* one in $F\ [x_1, \cdots, x_r]$.

Lemman 1. *There is an algorithmic procedure which permits to determine uniquely for any polynomial A in p_r, a set of integers $s_1, \cdots, s_r \geq 0$ and a set of polynomials A_I in p_0 for $I \in$ r verifying the following conditions:*

$(L1)_1$. *Modulo some linear combination of p_i over p_r we have*

$$p_{10}^{s_1} \cdots p_{r0}^{s_r} A \equiv \sum A_I x^I.$$

$(L1)_2$: *A_I are polynomials in p_0 with coefficients linear in those of A, considered as polynomial in p_r.*

$(L1)_3$: *$s_1, \cdots, s_r$ are the least integers ≥ 0 to make $(L1)_1$ and $(L1)_2$ possible.*

Proof. Considering both A and p_r as polynomials in x_r with coefficient in p_{r-1}, we get by division for some integer $s_r \geq 0$ taken to be least possible

$$p_{r0}^{s_r} A = Q_r p_r + R_{r-1}$$

with Q_r, R_{r-1} polynomials in P_r for which the degree of R_{r-1} in x_r is $< m_r$. Considering now p_{r-1} and R_{r-1} as polynomials in x_{r-1} with coefficients in $R[u_1, \cdots, u_d, x_1, \cdots, x_{r-2}, x_r]$, we get by division for some integer $s_{r-1} \geq 0$ taken to be least possible

$$p_{r-1,0}^{s_{r-1}} R_{r-1} = Q_{r-1} p_{r-1} + R_{r-2},$$

with Q_{r-1}, R_{r-2} polynomials in p_r for which the degrees of R_{r-2} in x_r and x_{r-1} are $< m_r$ and $< m_{r-1}$ respectively. Proceeding in this manner, we get successively

$$p_{r-2,0}^{s_{r-2}} R_{r-2} = Q_{r-2} p_{r-2} + R_{r-3},$$

$$\cdots\cdots$$

$$p_{10}^{s_1} R_1 = Q_1 p_1 + R_0,$$

with R_0 as a polynomial in pr, for which the degree in x_i is $< m_i$, $1 \leq i \leq r$. we may then write Ro as $\sum A_1 x^I$ and get the expression verifying all the conditions in the Lemma as required.

Lemma 2. *There is an algorithmic procedure which permits us to determine for any polynomial in some indeterminate y of the form*

$$A = A_0 y^m + A_1 y^{m-1} + \cdots + A_m$$

with each A_i *in* p_r *and* $A_0 \neq 0$ *in* K_r *expressions of the form*

$$HA = B + C_1 p_1 + \cdots + Crpr,$$

with

$$B = B_0 y^m + B_1 y^{m-1} + \cdots + B_m,$$

verifying the following conditions:

$(L2)_1$. *All* B_i *are normalized polynomials in* p_r *and Bo is a polynomial* $\neq o$ *in Po.*

$(L2)_2$. C_i *are all polynomials in* p_r $[y]$ *and H is one in Pr.*

$(L2)_3$. $H \not\equiv 0 \bmod (p_1, \cdots, p_r)$.

Any such polynomial B satisfying $(L2)_1$ *will then be said to be a normalized polynomial over* p_r *in the indeterminate* y.

Proof. By lemma 1, we have for some integers $s_1, \cdots, s_r \geq 0$ an expression of the form

$$P_{10}^{s_1} \cdots P_{r0}^{s_r} A = A'_0 y^m + A'_1 y^{m-1} + \cdots + A'_m \bmod (p_1, \cdots, p_r)$$

with all A'_i normalized in p_r and $A'_0 \neq 0$ in Kr. Suppose that A'_0 is free of $x_{i+1}, \cdots, x_r$ but not so for x_i. As $A'0$ has a degree $< m_i$ in x_i while p_i is irreducible in K_{i-1} and of degree m_i in x_i we find by the usual division algorithm polynomials h, k in p_i and A'_{00} in p_{i-1} such that

$$hA'_0 + kp_i = A'_{00},$$

in which h is $\neq 0$ in K_i and $A'_{00} \neq 0$ in K_{i-1}. We have then an expression of the form

$$hp_{10}^{s_1} \cdots p_{r0}^{s_r} A + kp_i y^m = A'_{00} y^m + hA'_1 y^{m-1} + \cdots + hA'm$$
$$\bmod (p_1, \cdots, p_r).$$

Applying Lemma 1 again to hA'_i and A'_{00} we get then some expression

$$h''A = A''_0 y^m + A''_1 y^{m-1} + \cdots + A''_m \bmod (p_1, \cdots, p_r),$$

for which all A''_i are normalized in p_r with $A''_0 \neq 0$ in K_r and not containing any x_j for $j \geq i$, and h'' is some polynomial in p_r which is $\neq 0$ in K_r. If A''_0 is free of all $x_1, \cdots, x_r$, we may take

$$B = A''_0 y^m + A''_1 y^{m-1} + \cdots + A''_m$$

as the polynomial B required or otherwise we proceed as before.

Lemma 3. *There is an algorithmic procedure which permits to factorize in K_r any polynomial in $p_r[y]$*

$$A = A_0 y^m + A_1 y^{m-1} + \cdots + A_m$$

with A_i in p_r and $A_0 \neq 0$ in K_r, $m \geq 2$. More precisely, it permits to find expression of the form

$$HA = A' + \sum C_i p_i,$$

with

$$A' = \overline{A_1} \cdots \overline{A_t},$$

verifying the following conditions:

$(L3)_1$. *Each $\overline{A_j}$ is a normalized polynomial in $p_r[y]$ and is irreducible in K_r.*

$(L3)_2$. *H is a polynomial in p_r which is $\neq 0$ in K_r.*

$(L3)_3$. *C_i are polynomials in $p_r[y]$.*

Proof. The method of Hermann in [4] permits us to give, in an algorithmic manner, a factorization of A into irreducible ones in K_r, so that after clearing of fractions we have an expression of the form

$$DA = B_1 \cdots B_t + \sum B'_i p_i,$$

with B_i, B'_i polynomials in $p_r[y]$, B_i irreducible in K_r and D a polynomial in p_r which is $\neq 0$ in K_r. Applying now Lemma 2 to

each B_i. we get then the expression required. Consider also the reference [15], p. 130.

IV. Proofs of the theorems

We are now in a position to give proofs of Theorems 1 – 3 quite simply as follows:

Proof of Theorem 1. For the defining system of equations (I) of the associated variety V, let us consider $f_1(u_1, \cdots, u_d, x_1)$ as a polynomial in x_1 with coefficients in $R[u_1, \cdots, u_d]$ and factorize f_1 into ones irreducible in $K = R(u_1, \cdots, u_d)$ by applying the algorithm in Lemma 3, with x_1 as y and 0 as r there. Take any such irreducible factor as $p_1(x_1)$. Let K_1 be the field obtained from K by adjoining x_1 defined by the equation

$$p_1(x_1) = 0$$

Now f_2, considered as a polynomial in the indeterminate x_2, cannot be identical with o in the field K_1, for otherwise the variety V would be of dimension $> d$, contrary to the determinancy hypothesis of our geometric statement. Applying Lemma 3 to f_2 with x_2 as y and 1 as r there, we get a certain polynomial f'_2 in $P_2 = R[u_1, \cdots, u_d, x_1, x_2]$ with an expression of the form

$$h_2 f_2 = f'_2 + C_{21} p_1,$$

in which f'_2, as a polynomial in the indeterminate x_2, is a product of normalized factors irreducible and $\neq 0$ in the field K_1. Take any such factor as $p_2(x_1, x_2)$. Let K_2 be the field obtained from K_1 by adjoining x_2 defined by the equation

$$p_2(x_1, x_2) = 0.$$

Then f_3 is a polynomial in the indeterminate x_3 over p_2 which cannot be identical with o. Applying Lemma 3, we get an expression of the form

$$h_3 f_3 = f'_3 + C_{31} p_1 + C_{32} p_2,$$

in which the polynomial f'_3 in x_3 is a product of normalized factors irreducible and $\neq 0$ in K_2. Take any factor as p_3 (x_1, x_2, x_3), adjoin x_3 to K_2 by the equation

$$p_3 (x_1, x_2, x_3) = 0,$$

and proceed further as before. In this manner, we get finally system of subvarieties V′ irreducible over R, each defined on K by systems of equations of the type

$$(\text{III}) \begin{cases} p_1 (x_1) = 0, \\ p_2 (x_1, x_2) = 0, \\ \qquad \cdots \\ p_r (x_1, x_2, \cdots, x_r) = 0, \end{cases}$$

verifying some obvious conditions as described in III.

It is easy to see that the collection of all such subvarieties exhausts the given variety V^*. In fact, consider an irreducible component in R of V^* with a generic point, say $(u_1, \cdots, u_d, \overline{x_1}, \cdots, \overline{x_r})$, in which $u_1, \cdots, u_d$, are independent indeterminates while $\overline{x_1}, \cdots, \overline{x_r}$ depend algebraically on them. As $(u_1, \cdots, u_d, \overline{x_1})$ should satisfy the system of equations (I), in particular the equation $f_1 = 0$, they should annul some one of the irreducible factors of f_1, say p_1 before. Now, $(u_1, \cdots, u_d, \overline{x_1}, \overline{x_2})$ should satisfy the other equations in the system (I) as well as the equation $p_1 (x_1) = 0$. It follows from the expression about f'_2 given above that they should also satify the e-

quation $f'_2=0$ and hence should annul one of the irreducible factors of f'_2, say $P_2(x_1, x_2)$ above. Proceeding in the same manner we see that $(u_1, \cdots, u_d, \overline{x_1}, \cdots, \overline{x_r})$ should satisfy a system of equations of the type (Ⅲ) and hence is a generic point of an irreducible subvariety among the collection found above. This completes the proof of the theorem.

Proofs of Theorems 2 *and* 3.

Theorem 2 follows immediately from the algorithm in Lemma 1. The truth of Theorem 3 is quite clear from Theorems 1 and 2.

Final Remark. Our algorithm for the mechanization of theorem-proving in elementary geometry involves mainly such polynomial manipulations as arithmetic operations and simple eliminations, which were all originated and quite developed in 12-14c. Chinese mathematics, of. e.g. the book of late Chien [1] for the explanations. In fact, the algebrization of geometrical problems and systematic method of their solutions by algebraic tools were some of the main achievements of Chinese mathematicians at that time, much earlier than the appearance of analytic geometry in 17c.

Added in Proof. (Dec. 1977).

The same principle has been applied to the mechanization of theorem-proving in elementary differential geometry with the aid of Ritt's theory of differential algebra, of which the details will be given later.

Bibliography

[1] Chien, Baozong. *The history of Chinese mathematics*, Science Press,

1964.

〔2〕 Cohen, P. J. Decision procedures for real and p - adic fields, *Comm. Pure Appl. Math*, 22 (1969), 131—152.

〔3〕 Gröbner, W. *Moderne algebraische Geometrie* (1949), Wien.

〔4〕 Hermann, G. Die Frage der endlich vielen Schritte in der Theorie der Polynomideale, *Math. Ann.*, 95 (1926), 736～788.

〔5〕 Hilbert, D. *Grundlagen der Geometrie*.

〔6〕 McCharen, J. D. *et al*. Problems and experiments for and with automated theorem - proving programs, *IEEE Tarns. on Computers*, C - 25 (1976), 773 - 782.

〔7〕 Reiter, R. A semantically guided deductive system for automatic theorem proving, *IEEE Trans. on Computers*, C - 25 (1976), 328 - 334.

〔8〕 Ritt, R. F. *Differential Algebra* (1950).

〔9〕 Robinson, A. *Introduction to Model Theory and to the Metamathematics of Algebra* (1963), Amsterdam.

〔10〕 Robinson, A. Algorithms in algebra, in model theory and algebra, *Lect. Notes in Math*. No. 498 (1975), 14 - 40.

〔11〕 Robinson, A. A decision method for elementary algebra and geometry - revisited, *Proc. of Tarski Symposium* (1974), 139 - 152.

〔12〕 Seidenberg, A. A new decision method for elementary algebra, *Annals of Math*, 60 (1954), 365 - 374.

〔13〕 Tarski, A. What is elementary geometry? *The Axiomatic Method with. Special Reference to Geometry and Physics* (1959), 16 - 29, Amsterdam.

〔14〕 Tarski, A. and Mckinsey, J. C. C. *A Decision Method for Elementary Algebra and Geometry*, 2nd ed., Berkeley and Los Angeles (1948 - 1951).

〔15〕 Waerden, Van der. *Moderne Algebra*, Bd. 1 (1930).

〔16〕 Waerden, Van der. *Einführung in die algebraischen Geometrie* (1945).

Toward mechanization of geometry—some comments on Hilbert's "Grundlagen der Geometrie" *

The great merit of Hilbert's classic "Grundlagen der Geometrie" of 1899 is universally recognized as being represenative for axiomatization of mathematics, laying in particular a rigorous foundation of the euclidean geometry. However, another great merit (perhaps greater in the opinion of the present author) of this classic seems hardly to be noticed up to the present. In fact, this classic is also representative for the mechanization of geometry, showing clearly at the same time the way to achieve it. The present paper has the object of trying to clarify these points.

First of all let us remark that in the statement of theorems (or even axioms) there are usually some implicit assumptions about *gene* – ricity or *non-degeneracy* of figures involved without which the theorems may be meaningless or even fall into fallacies. The following simple examples may serve as illustrations.

Ex. 1 The three altitudes of a triangle are concurrent.

It is implicitely assumed here that the triangle in question

---

* A cta Math. Scientia, 2 (1982) 125—138

should be *generic* in the sense that it does not degenerate into one with vertices collinear or with two vertices coincident. In the first case the theorem is not true and in the second case it is meaningless.

Ex. 2 The opposite sides of a parallelogram are congruent.

The theorem will not be true if the parallelogram in question degenerates into one with vertices lying on the same line.

Ex. 3 Desargues Theorem. If the three pairs of corresponding sides of two triangles are all parallel to each other, then the three lines joining the corresponding vertices of these triangles are either concurrent or are parallel to each other.

The theorem will not be true if the two triangles in question degenerate into ones with collinear vertices.

Ex. 4 Desargues Theorem. If the three lines joining the corresponding vertices of two triangles are either concurront or parallel, and two pairs of corresponding sides of these triangles are both parallel to each other, then the same will be true for the third pair of corresponding sides.

Again the theorem will not be true if one of the pairs of corresponding sides parallel to each other degenerates into a pair lying on the same line.

In view of these examples we see that theorems of elementary geometry are usually true only in the *generic* or *non − degenerate* case which are implicitely assumed as hypothesis but usually not clearly expressed in the statements of the theorems. In each degenerate case we have to investigate separately whether the theorem is meaningful or not and if it is so whether the theorem remains hold true or not.

Now to prove theorems in the usual euclidean fashion one should incessantly make resort to previously proved theorems considered to be already known. As these known theorems are only true under certain non-degeneracy conditions one should, each time when these theorems are to be applied, verify whether these non-degeneracy conditions are observed or not. One should consider different cases to deprive off one by one each of these degeneracy situations. This renders the proof of a theorem very cumbersome, the more so because such non-degeneracy conditions of theorems to be applied are usually not clearly stated. In fact, even for a theorem of moderate complexity, it would be quite impossible to take care of all these non – degenerate cases occured in the known theorems to be applied in order to make the proof meeting the usual standard of rigor. A scrutiny of the proofs of theorems concerning the establishment of rules of number systems in a Desarguesian geometry, in which Desargues theorem should be applied over and over, as described in the classic of Hilbert, may well illustrate this point.

On the other hand Hilbert, in his classic, after laying down the foundation for the algebraization and coordinatization of a geometry involving Desargues theorem and Pascal theorem, has stated a theorem (numbered Th. 62 in the later editions), which is in essence equivalent to a mechanized procedure for the proofs of a certain kind of theorems. Let us call for short a plane geometry to be *pascalian* if the planar axioms of incidence, the axioms of order, the axiom of parallelism in the strengthened form, as well as the so-called Pascal theorem all hold true. Then the original statement of the theorem in question runs somewhat

as follows.

Th. 62 Each *pure intersection-point theorem*, in a Pascalian geometry, if true, can always be proved by the aid of suitably constructing auxiliary points and lines, as consequence of a combination of a finite number of Pascal configurations.

For the meaning of pure intersection-point theorem (abb. PIP-Th) Hilbert has given two explanations which are in fact not equivalent. For the first one Hilbert defined a PIP-Th as one in which only incidence of points and lines as well as parallelism of lines are involved. Then, he argued that every such PIP-Th. may be described in more details in the following form:

Select first a system of finite number of points and lines arbitrarily. Then successively in a definite prescribed manner draw some parallels to some of the lines get, choose some points on some of the lines get, and draw some lines through some of the points get. If, by constructing in this prescribed manner the joining lines, intersection points as well as parallels through points already constructed, one arrives finally to a finite set of lines, then the theorem will assert that these lines will be either concurrent or parallel to each other.

To distinguish between these two concepts of PIP-Th, in fact not equivalent, we shall call the later one the PIP-Th *of constructive type*. Now the explanations given just before the statement of Theorem 62 in the classic furnish an idea of a proof of the following theorem, which may be considered as an alternative version of Th. 62 and will be called

Mechanization Theorem of Hilbert. In a planar Pascalian geometry there is a mechanical procudure which permits to prove

or disprove in a finite number of steps any pure intersection – point theorem of constructive type under certain subsidiary non-degeneracy conditions also generated mechanically during the procedure.

As the intended proof of Theorem 62 or the above Mechanization Theorem as given in Hilbert's classic is not only vague but also somewhat inexact in various respects, we shall rewrite the proof in what follows which consists in giving a mechanical procedure meeting the purpose as described in the Mechanization Theorem.

First of all we shall take in the plane a coordinate system which may be chosen arbitrarily. To save the labor of computations we may choose, if we want, the coordinate system in a convenient manner which is however immaterial to the reasonings to follow. The points and lines involved in any pure intersection-point theorem will then be represented by number pairs or linear equations. To fix the ideas, we shall avoid the use of linear equations for lines and restrict ourselves to considerations of points or number pairs alone as described below. As a point may be either chosen in an arbitrary manner or constructed in a definite manner from prescribed geometric conditions, we shall distinguish two types of coordinates: parametric ones and geometrically-bounded ones, which will be denoted by u_i and x_j respectively. They will be denoted by a_k if no distinction for u or x is necessary . We shall now represent the points occuring successively in the PIP-Th. of constructive type to be proved by number pairs $(a_i, \ a_j)$ one after the other as follows.

First let us remark that the construction of a point occuring

in the theorem in question will be one of the following 10 types:

1. A point is arbitrarily given or chosen.

The point will then be represented as (u_i, u_j) with u_i, u_j as parameters.

2. A line is arbitrarily given or chosen.

Instead of representing the line by a linear equation of current coordinates, we shall take arbitrarily two points represented by (u_i, u_j), (u_k, u_l) respectively with the line in question as their joining line.

3. Construct an arbitrary line through a point (a_i, a_j) already constructed.

As in 2, we shall take an arbitrary point (u_k, u_l) and represent the line as the joining line of this point and the point (a_i, a_j).

4. Construct the joining line of two points already constructed.

As the line has been determined by the two points thereon, this construction is no more necessary.

5. Choose a point arbitrarily from a line already constructed.

If the line is determined by two points (a_i, a_j), (a_k, a_l) already constructed, then the arbitrary point chosen thereon may be either represented by (u_r, x_s) or (x_s, u_r) satisfying the following equation:

$$(a_j - a_i)u_r - (a_i - a_k)x_s + a_ia_l - a_ja_k = 0,$$

or

$$(a_j - a_l)x_s - (a_i - a_k)u_r + a_ia_i - a_ja_k = 0.$$

6. Construct arbitrarily a parallel to a line already constructed. As the line is determined by two points thereon already cons

tructed say (a_i, a_j), (a_k, a_l), the parallel to be constructed will be determined as follows. First take an arbitrary point (u_m, u_n), and then a point (u_r, x_s) or (x_s, u_r) such that the joining line of this point to the point (u_m, u_n) will be parallel to the line determined by
(a_i, a_j) and (a_k, a_l), so that x_s will satisfy the following equation:

$$(a_k - a_i)(x_s - u_n) - (a_l - a_j)(u_r - u_m) = 0,$$

or

$$(a_l - a_j)(x_s - u_m) - (a_k - a_i)(u_r - u_n) = 0.$$

7. Construct a line through a point (a_m, a_n) already constructed parallel to a line already constructed.

Let the line already constructed be determined by the points (a_i, a_j) and (a_k, a_l), then the line to be constructed will be taken to be one determined by (a_m, a_n) and a further point (u_r, x_i) or (x_s, u_r) satisfying the following equation:

$$(a_k - a_i)(x_s - a_n) - (a_l - a_j)(u_r - a_m) = 0,$$

or

$$(a_l - a_j)(x_s - x_m) - (a_k - a_i)(u_r - a_n) = 0.$$

8. Construct the intersection-point of two intersecting lines already constructed.

Let the two lines be determined respectively by pairs of points (a_i, a_j), (a_k, a_l) and (a_p, a_q), (a_r, a_s) already constucted. Then the intersection-point will be taken to be (x_g, x_h), satisfying the following system of equations:

$$(a_j - a_l)x_g - (a_i - a_k)x_h + a_i a_l - a_j a_k = 0,$$

and

$$(a_q - a_s)x_g - (a_p - a_r)x_h + a_p a_s - a_q a_r = 0.$$

9. Construct the intersection-point of a line already constructed and a line through a point (a_m, a_n) already constructed and parallel to a second line already constructed.

Let the two lines already constructed be determined respectively by (a_i, a_j), (a_k, a_l) and (a_p, a_q) (a_r, a_s). Represent the point to be constructed by (x_g, x_h), then x_g, x_h will satisfy the following system of equations:

$$(a_r - a_p)(x_h - a_n) - (a_s - a_q)(x_g - a_m) = 0,$$

$$(a_j - a_l)x_g - (a_i - a_k)x_h + a_i a_l - a_j a_k = 0.$$

10. Construct the intersection-point of two lines passing through each of two points already Constructed and parallel respectively to each of two lines already constructed.

This is similar to 8 and 9 and may be similarly treated.

Of course we may reduce the constructions 9 and 10 to the previous ones by introducing more coordinates u or x.

We are now in a position to describe the mechanical procedure to be followed. As the intersection-point theorem to be proved is of constructive type, the points and lines as described in the theorem will occur one after the other in a definite order of succession. It follows that the coordinates of the points involved (with lines replaced by two points thereon), whether parametric or geometrically-bounded, can be arranged in a definite order in accordance with their ordering of appearance in the construction as follows:

$$u_1 \prec u_2 \prec \cdots\cdots \prec u_m,$$

$$x_1 \prec x_2 \prec \cdots\cdots \prec x_n.$$

In particular, whenever a new point is introduced by the construction 8, 9 or 10, its coordinates attributed will be two x's in succession, say (x_g, x_{g+1}).

Let us introduce now some sets as follows.

A set of parametric coordinates

$$U = \{u_1, \cdots, u_M\}.$$

A set of geometrically – bounded coordinates

$$X = \{x_1, \cdots, x_N\}.$$

A set of *degeneracy – polynomials*

$$\Delta = \{D_1, \cdots, D_R\}.$$

with each D a polynomial$\not\equiv 0$ in $u_1, \cdots_M$ alone.

A set of *solutions*

$$S = \left\{\frac{Q_1}{P_1}, \cdots, \frac{Q_N}{P_N}\right\},$$

with all P_i'S, Q_i'S polynomials in $u_1, \cdots, u_M$ alone and each P_i a nonzero power product of the D'S in Δ.

The meaning of the sets U and X are already clear. The meaning of the sets Δ and S is this:

Under the non-degeneracy conditions

$$D_1 \neq 0, \cdots, D_R \neq 0 \qquad (D_i \text{ in } \Delta)$$

the coordinates x_i are given by

$$x_1 = \frac{Q_1}{P_1}, \cdots, x_N = \frac{Q_N}{P_N}.$$

with x_i in X, Q_i/P_i in S.

We begin now by setting all the sets X, Δ, and S to be empty ones, U to be $\{u_1, \cdots, u_m\}$, and proceed by enlarging these sets in following in steps the successive constructions as described in the theorem to be proved in question.

Suppose that we have proceeded to a certain step of construction but not yet finished. We distinguish two cases according as the next step of construction is one of the types 1-7 or of the types $8-10$. In the first case we get a single equation

$$A_{N+1}x_{N+1}+B_{N+1}=0$$

with a new bounded coordinate x_{N+1} and with A_{N+1}, B_{N+1} both polynomials in u_1, $\cdots$, u_M and x_1, $\cdots$, x_N already occured in the sets U and X. Replace now x_i's in A_{N+1} and B_{N+1} by Q_i/P_i given already in S, with $P_i \neq 0$ due to the already in troduced ono-degeneracy-condition S $D_i \neq 0$ with D_i in Δ already determined. Let the new fractional expressions thus obtained in u_1, $\cdots$, u_N alone be denoted by $A^*_{N+1}$and B^*_{N+1} respectively so that the equation in x_{N+1} becomes

$$A^*_{N+1}X_{N+1}+B^*_{N+1}=0.$$

Several. possibilities may then occur.

If$A^*_{N+1} \not\equiv 0$ then we write the numerator of A^*_{N+1} as D_{R+1} and put D_{R+1} into the set Δ to turn it into a new Δ, and introduce as a new non-degeneracy condition $D_{R+1} \neq 0$. Next we solve for x_{N+1} in the form

$$x_{N+1}=-B^*_{N+1}/A^*_{N+1}=Q_{N+1}/P_{N+1},$$

and put x_{N+1} into X and Q_{N+1}/P_{N+1} into S to turn X and S into new set X and new solution set S. We note that P_{N+1} is again a power product of the polynomials D'_jS in the new set Δ.

We proceed then to the next step of construction if there remains any.

If instead $A^*_{N+1} \equiv 0$, but $B^*_{N+1} \not\equiv 0$, then we set again $D_{R+1}=$ numerator of B^*_{N+1} and put it into Δ. The hypothesis of

the theorem in question is now itself contradictory at least under the subsidiary conditions

$$D_j \neq 0$$

for D_j in the new Δ. The whole procedure will then be stopped.

If finally both A^*_{N+1} and $B^*_{N+1} \equiv 0$, then x_{N+1} undergoes in reality no restrictions and we may introduce a new parameter u_{M+1} and set

$$x_{N+1} = u_{M+1} = Q_{N+1}/P_{N+1}$$

(with $Q_{N+1} = u_{M+1}$, $P_{N+1} = 1$, the same in what follows). We put then u_{M+1} into U, x_{N+1} into X, and Q_{N+1}/P_{N+1} into S with Δ unchanged. With this new system of sets U, X, Δ and S we then proceed to the next step of construction if there remains any.

This ends the treatment in the first case.

Suppose now we are in the second case of a construction of the types 8, 9, or 10. For any such type we will get a system of two equations say

$$A_{11}x_{N+1} + A_{12}x_{N+2} + B_1 = 0,$$

$$A_{21}x_{N+1} + A_{22}x_{N+2} + B_2 = 0,$$

with two new bounded coordinates x_{N+1}, x_{N+2} introduced atthe same time. The polynomials $A'S$ and $B'S$ are allin the variables $u_1, \cdots, u_n$ of U and $x_1, \cdots, x_N$ in X. Replace now these $x_i'S$ by Q_i/P_i in S for $i = 1, \cdots, N$ we get for the $A'S$ and $B'S$ some fractional expressions in the $u'S$ of U alone to be denoted by $A^*{}'S$ and $B^*{}'S$. In this way the equations above become then the following ones:

$$A^*_{11}x_{N+1} + A^*_{12}x_{N+2} + B^*_1 = 0,$$

$$A_{21}^* x_{N+1} + A_{22}^* x_{N+2} + B_2^* = 0,$$

Various possibilities may now occur.

First suppose that the determinant of the coefficients of these equations in x_{N+1}, x_{N+2} is not identically zero:

$$E = \begin{vmatrix} A_{11}^* & A_{12}^* \\ A_{21}^* & A_{22}^* \end{vmatrix} \not\equiv 0.$$

In this case let us express E in the form of a fraction of which the denominator is necessarily a power-product of D_j'S already present in Δ. We set now the numerator of E to be a new D_{R+1} and put it into Δ to enlarge it into a new one still denoted by Δ. Solve the two equations above we get then two expressions of the form

$$x_{N+1} = Q_{N+1}/P_{N+1}, \quad x_{N+2} = Q_{N+2}/P_{N+2},$$

in which both P_{N+1} and P_{N+2} are power-products of $D_j S'$ in the new Δ Wc put now x_{N+1}, x_{N+2} into X, and $\frac{Q_{N+1}}{P_{N+1}}$, $\frac{Q_{N+2}}{P_{N+2}}$ in to S, denote the sets thus enlarged still by X and S, and then proceed to the next step of construction if there remains any.

Suppose now E is identically 0 but not all A^*'S are so. In this case there will exist polynomialsa_1, a_2 not both 0 such that

$$a_1 A_{11}^* + a_2 A_{21}^* \equiv 0,$$

$$a_1 A_{12}^* + a_2 A_{22}^* \equiv 0.$$

From the two equations above we get

$$a_1 B_1^* + a_2 B_2^* = 0.$$

If $a_1 B_1^* + a_2 B_2^* \not\equiv 0$, then we set

$$D_{R+1} = \text{numerator of } a_1 B_1^* + a_2 B_2^*$$

and put it into Δ. In this case the hypothesis of the theorem in

question is contradictory in itself under the non-degeneracy conditions

$$D_j \neq 0,$$

with D_j in the new Δ. We stop then and proceed no more.

In case $a_1 B_1^* + a_2 B_2^*$ is $\equiv 0$, the two equations will reduce to a single one, say the first one to fix the ideas. Not both A_{11}^* and A_{12}^* can be $\equiv 0$, say $A_{12}^* \not\equiv 0$ to fix the ideas.
Set then

$$D_{R+1} = \text{Numerator of } A_{12}^*$$

and put it into Δ. Introduce also a new u_{M+1} to be put in U, set

$$x_{N+1} = u_{M+1} = Q_{N+1}/P_{N+1}$$

and solve for x_{N+2} in the form

$$x_{N+2} = Q_{N+2}/P_{N+2}$$

with P_{N+2} a power-product of D_j'S in the new Δ. We put now x_{N+1}, x_{N+2} into X and Q_{N+1}/P_{N+1}, Q_{N+2}/P_{N+2} into S. All the sets Δ, U, X, S thus enlarged will again be denoted by the same letters and then proceed to the next step of construction if there remains any.

Finally let us suppose that all the A^*'S are identically 0. If at least one of B^*, B_2^* is not identically 0 then we set the numerators of these non-zero B^*'S as new D_{R+1} or D_{R+1} and D_{R+2}, put this one or both into Δ with the new Δ still denoted by the same letter. We stop then the whole procedure with the conclusion that the hypothesis of the theorem is in contradiction under the non-degeneracy conditions $D_j \neq 0$ with D_j in the new Δ.

If finally both B_1^* and B_2^* are identically zero, then we introduce two new u_{M+1}, u_{M+2} to be put into U, put x_{N+1}, x_{N+2}, into X, and set

$$x_{N+1} = u_{M+1} = Q_{N+1}/P_{N+1},$$
$$x_{N+2} = u_{M+2} = Q_{N+2}/P_{N+2},$$

with Q_{N+1}/P_{N+1}, Q_{N+2}/P_{N+2} to be put into S. The new sets U, etc. will then again be denoted by the same letters. We proceed now to the next step of construction if there remains any.

This procedure will be stopped with a final system of sets:

$$U = \{u_1, \cdots, u_M\},$$
$$X = \{x_1, \cdots, x_N\},$$
$$\Delta = \{D_1, \cdots, D_R\},$$

and

$$S = \{\frac{Q_1}{P_1}, \cdots, \frac{Q_N}{P_N}\}.$$

Now two cases may occur.

Case 1 The hypothesis of theorem in question is contradictory under the subsidiary non-degeneracy conditions

$$D_j \neq 0 \quad (D_j \text{ in } \Delta).$$

In this case we have already achieved our aim of proof.

Case 2 In this contrary case we have to consider the conclusion of the theorem in question which amounts to say that

$$G = 0,$$

where G is a certain polynomial in the variables $u_1, \cdots, u_m$ and $x_1, \cdots, x_n$. Now N is necessarily $= n$ and $M \geq m$ in the present case and we may thus put

$$x_i = Q_i/P_i, \quad i = 1, \cdots, n,$$

and substitute these expressions into G. The polynomial G be-

comes then a fraction with denominator a power - product in the D_j*'S* occuring in Δ. Under the non - degeneracy conditions

$$D_j \neq 0 \quad (D_j \text{ in } \Delta),$$

we verify by usual computations whether G is identically zero or not and arrive at the two possible conclusions below.

1° G becomes identically zero so that the theorem is true under the above non - degeneracy conditions. The whole procedure constitutes then an actual proof of this theorem with restrictions clearly exhibited by the above non - degeneracy conditions.

2° G is not identically zero so that the theorem is not true at least under the above non - degeneracy conditions.

This terminates the whole mechanical procedure with precise conclusions under precise subsidiary conditions also mechanically generated.

The mechanization theorem of Hilbert is thus completely proved.

As illustrations of the above mechanical proving procedure for intersection - point theorems let us consider some examples shown below.

Ex. 5 Consider the Desargues theorem of Ex. 3 stated in the following form: Let ABC and $A'B'C'$ be two triangles with three pairs of corresponding sides mutually parallel to each other. Let the line $l_1 = AA'$ and $l_2 = BB'$ meet at a point $\mathbf{0}$. Then CC' should also pass through $\mathbf{0}$.

Let us turn the statement of the theorem into a constructive form as follows. To simplify the calculations we shall take l_1, l_2 as the two coordinate axes. We shall take first two arbitrary points A, A' on l_1 with coordinates

$$A=(u_1,0),\ A'=(u_2,0).$$

Then take an arbitrary point B on l_2 and an arbitrary point C on the plane with coordinates

$$B=(0,\ u_3),\ C=(u_4,\ u_5).$$

Draw now through A' a parallel of AB meeting l_2 in B' and then through A' and B' draw parallels to AC and BC respectively to meet together at C'. The coordinates of the new points will be taken to be

$$B'=\ (0,\ x_1),\ \ C'=\ (x_2,\ x_3).$$

Then the theorem asserts that the points O, C, C' are collinear.

Now the hypothesis of the theorem consists of the following equations

$$A'B' // AB \Leftrightarrow u_1x_1-u_2x_3=0,$$

$$A'C' // AC \Leftrightarrow (u_4-u_1)x_3-u_5(x_2-u_2)=0,$$

$$B'C' // BC \Leftrightarrow u_4(x_3-x_1)-(u_5-u_3)x_2=0.$$

The conclusion of the theorem is given by:

$$O,\ C,\ C' \text{ are collinear} \Leftrightarrow G=0,$$

with

$$G=u_4x_3-u_5x_2.$$

Following the mechanical procedure indicated above we get

$$U=\{u_1,\ u_2,\ u_3,\ u_4,\ u_5\},$$

$$X=\{x_1,\ x_2,\ x_3\},$$

$$\Delta=\{u_1,\ u_1u_3-u_1u_5-u_3u_4\},$$

$$S=\left\{\frac{u_2u_3}{u_1},\ \frac{u_2u_4}{u_1},\ \frac{u_2u_5}{u_1}\right\}.$$

If follows that the non $-$ degeneracy conditions to be observed are

$$u_1 \neq 0,$$

and

$$u_1 u_3 - u_1 u_5 - u_3 u_4 \neq 0.$$

Under these conditions we verify that

$$G = u_4 x_3 - u_5 x_2 = u_4 \cdot \frac{u_2 u_5}{u_1} - u_5 \cdot \frac{u_2 u_4}{u_1} = 0.$$

so that the Desargues theorem in question is true in this non-degenerate case.

The first degeneracy case $u_1 = 0$ means that A coincides with O. It is of no geometrical interest and may be left out of consideration at all. If we want we may take $A = O$ as one of the hypothesis of the theorem and proceed in the same mechanical manner as before with the result that the theorem will then be trivially true under however some further non-degeneracy conditions.

The second degeneracy condition

$$u_1 u_3 - u_1 u_5 - u_3 u_4 = 0.$$

means that the points A, B, C are collinear. Take this as one of the hypothesis we verify as before that the theorem will then be no more true at least under certain further non-degeneracy conditions. This accounts for the fallacies of the Desargues theorem as indicated in Ex. 3. The same may be said about the other examples 1, 2 and 4. What is important for us is that the degeneracy conditions which may cause the fallacies of theorems present *automatically* during a mechanical procedure and may be treated alternatively and systematically also in a mechanical way which is actually impossible for the usual euclidean fashion proofs.

Ex. 6 To give a less trivial example let us consider the general Pascal theorem in projectivo-geometry. Let us call for

short a hexagon to be *Pascalian* if the three pairs of opposite sides intersect in collinear points. Now the necessary and sufficient condition for a hexagon to be Pascalian is that the six vertices of the hexagon should lie on the same conic (or *co* − *conic* for short). It follows that if a hexagon $A_1A_2A_3A_4A_5A_6$ is Pascalian, then, the six vertices being co − conic, any hexagon $A_{i1}A_{i2}A_{i3}A_{i4}A_{i5}A_{i6}$ arising from reordering of the vertives, is also pascalian. In this way we may announce theorems in the form of pure intersection − point theorems in a Pascalian geometry with the mention of the notion of conics completely avoided. The intersection − point theorems thus arrived may be divided into various types and, to fix the ideas, let us consider for example the following one.

If the hexagon $A_1A_2A_3A_4A_5A_6$ is Pascalian then so is $A_1A_4A_3A_2A_5A_6$.

In more details, the theorem states that:

If the points of intersection

$$P = A_1A_2 \wedge A_4A_5,\ Q = A_2A_3 \wedge A_5A_6,\ R = A_3A_4 \wedge A_6A_1$$

are collinear, then so are the points of intersection

$$P' = A_1A_4 \wedge A_2A_5,\ Q'' = A_3A_4 \wedge A_5A_6,\ R' = A_2A_3 \wedge A_6A_1$$

(Here $\wedge$ stands for intersection).

For the proof let us first turn the above statement into one of constructive type as follows.

Take first an arbitrary point A_6 and then two arbitrary lines l_1, l_2 through A_6. For the mere sake of simplifying the calculations we shall take A_6 as the origin 0 and the two lines as the coordinate axes.

Take on l_1 an arbitrary point $A_1 = (u_1, 0)$.

Through A_1 construct an arbitrary line and take thereon an arbitrary point $A_2 = (u_2, u_3)$.

Through A_2 construct an arbitrary line and take thereon an arbitrary point $A_3 = (u_4, u_5)$.

Through A_3 construct an arbitrary line and take thereon an arbitrary point $A_4 = (u_6, u_7)$.

Let the line A_3A_4 meet l_1 in the point $R = (x_1, 0)$.

Let the line A_2A_3 meet l_2 in the point $Q = (0, x_2)$.

Let the line A_3A_4 meet l_2 in the point $Q' = (0, x_3)$.

Let the line A_2A_3 meet l_1 in the point $R' = (x_4, 0)$.

Join the line QR and let it meet A_1A_2 in the point $P = (x_5, x_6)$.

Join PA_4 and let it meet l_2 in the point $A_5 = (0, x_7)$.

Join A_1A_4 as well as A_2A_5 meeting in the point $P' = (x_8, x_9)$.

With these hypothesis the conclusion is now: P', Q', R' are collinear.

Now the hypothesis of the theorem in question reads as follows:

R lies on $A_3A_4 \Leftrightarrow (u_5 - u_7)x_1 + u_4u_7 - u_5u_6 = 0$,

Q lies on $A_2A_3 \Leftrightarrow (u_2 - u_4)x_2 - u_2u_5 + u_3u_4 = 0$,

Q' lies on $A_3A_4 \Leftrightarrow (u_4 - u_6)x_3 - u_4u_7 + u_5u_6 = 0$,

R' lies on $A_2A_3 \Leftrightarrow (u_3 - u_5)x_4 + u_2u_5 - u_3u_4 = 0$,

P lies on $A_1A_2 \Leftrightarrow u_3x_5 + (u_1 - u_2)x_6 - u_1u_3 = 0$,

P lies on $QR \Leftrightarrow x_2x_5 + x_1x_6 - x_1x_2 = 0$,

$A_5 = PA_4 \wedge l_2 \Leftrightarrow (u_6 - x_5)x_7 + u_7x_5 - u_6x_6 = 0$,

P' lies on $A_1A_4 \Leftrightarrow u_7x_8 + (u_1 - u_6)x_9 - u_1u_7 = 0$,

P' lies on $A_2A_5 \Leftrightarrow (u_3 - x_7)x_8 - u_2x_9 + u_2x_7 = 0$.

The conclusion becomes:

P', Q, R' are collinear $\Leftrightarrow G \equiv x_3x_8 + x_4x_9 - x_3x_4 = 0$.

Fowllowing the mechanical procedure given in the Mechanization Theorem of Hilbert we see that the theorem is true after a long and tedious but easy and mecahnical computations under certain subsidiary non- degeneracy conditions uninteresting to be explicitely given. The computations constitute then automatically a proof of the theorem in the generic or non - degeneracy case.

Ex.7 Let us consider the previous example again with however the theorem in question not turned into constructive type. Thus, let us take A_6A_1, A_6A_5 still as the coordinate axes but with coordinates of the various points as follows.

$A_1 = (u_1, 0)$, $A_2 = (u_2, u_3)$,

$A_3 = (u_4, u_5)$, $A_5 = (0, u_6)$,

$A_4 = (u_7, x_1)$, $Q = (0, x_2)$,

$R = (x_3, 0)$, $R' = (x_4, 0)$,

$P = (x_5, x_6)$, $Q' = (0, x_7)$,

$P' = (x_8, x_9)$.

The hypothesis of the theorem becomes then

A_4, A_5, P are collinear $\Leftrightarrow (x_1 - u_6)x_5 - u_7x_6 + u_6u_7 = 0$,

P, Q, R are collinear $\Leftrightarrow x_2x_5 + x_3x_6 - x_2x_3 = 0$,

A_3, A_4, R are collinear $\Leftrightarrow (x_1 - u_5)x_3 - u_4x_1 + u_5u_7 = 0$,

A_1, A_2, P are collinear $\Leftrightarrow u_3x_5 + (u_1 - u_2)x_6 - u_1u_3 = 0$,

A_2, A_3, Q are collinear $\Leftrightarrow (u_2 - u_4)x_2 - u_2u_5 + u_3u_4 = 0$,

etc.

From these we get an equation of the form

$$Ax_1^2 + Bx_1 + C = 0,$$

with A, B, C all polynomials in the u's alone and A not identically zero. It follows that under the non-degeneracy condition $A \neq 0$ the previous mechanical procedure does not work.

The last example shows clearly that we should distinguish between two types of pure intersection-point theorems, those of constructive type and those not, and that the idea of the mechanical procedure as indicated by Hilbert in his classic works only for theorems of constructive type. For pure intersection-point theorems which cannot be turned into constructive form or for theorems which involve non-linear equations as are usual for various kinds of geometries, we have to device other mechanical procedures to give mechanical proofs. Now such mechanical procedures do exist for a large class of theorems so far no order relations are involved and the procedures are even *feasible* in the sense that quite difficult theorems may be proved in this manner on a computer of moderate size in a reasonable period of time. The same can even be done for (local) differential geometry. We shall not enter into this more which has been sketched in some original papers of the present author. A book with details is now also in preparation.

Some remarks on mechanical theorem-proving in elementary geometry*

Dedicated to Professor Lee Kwok-ping (Li Guoping) on the Occasion of his 50th Year of Educational and Scientific Work

In a previous paper (cf. [3]) we have pointed out that almost all theorems in ordinary geometry (i. e. ordinary euclidean geometry) are only generically true, or true only under some non-degeneracy) conditions usually not explicitely described in the statement of the theorem. This fact is quite fatal to make the usual euclidean fashion of proving theorems rigorous as one believes to be so, since it is difficult to state clearly the non-degeneracy conditions to be observed and to verify that the previous theorems to be applied in the proof of the theorem in question do not fall into the degeneracy cases under which the previous theorems to be applied might not be true. On the contrary our mechanical method of theorem-proving in geometry permits to discover automatically the non-degeneracy conditions to be observed and offer the means of dealing with such degeneracy conditions cases in systematic mechanical way. It is in fact this crucial point which is

---

*A cta Mata. Scientia 3 (1983) 357-360.

responsible for the high efficiency of our method in proving quite non-trivial theorems without the defect pointed out above. In the paper [3] we have already indicate some examples to illustrate the above point. Let us take one of them, say the *Desargues* theorem, and try to explain it in some more details as follows.

The *Desargues* theorem in question says that if the corresponding sides of two triangles $A_1A_3A_5$ and $A_2A_4A_6$ are natually parallel, then the lines joining corresponding vertices will be either concurrent or all parallel to each other, and conversely.

To fix the ideas, we shall consider only one of the four parts of this theorem. Thus, let us suppose that A_3A_4 and A_1A_2 meet and take an oblique coordinate system with A_1A_2, A_3A_4 the two axis meeting at the origin. Let

$$A_1=(x_1,0), \qquad A_2=(x_2,0),$$
$$A_3=(0,x_3), \qquad A_4=(0,x_4),$$
$$A_5=(x_5,x_6), \qquad A_6=(x_7,x_8).$$

Then the hypothesis are:

(1) $A_1A_3 /\!/ A_2A_4$, or

$$f_1 \equiv x_1x_4 - x_2x_3 = 0,$$

(2) $A_1A_5 /\!/ A_2A_6$, or

$$f_2 \equiv (x_5 - x_1)x_8 - x_6(x_7 - x_2) = 0,$$

(3) $A_3A_5 /\!/ A_4A_6$, or

$$f_3 \equiv x_5(x_8 - x_4) - (x_6 - x_3)x_7 = 0.$$

The conclusion is:

A_5A_6passes through the origin, or

$$g \equiv x_5x_8 - x_6x_7 = 0.$$

The theorem says that the conclusion $g=0$ is a consequence

of the hypothesis $f_1 = 0$, $f_2 = 0$ and $f_3 = 0$, but only generically so. It is easy to be tempted to prove the theorem in showing g or some of its power lies in the ideal (f_1, f_2, f_3) generated by f_1, f_2, f_3. This is surely not so, as is easily verified. In fact, the theorem is only generitcally true under certain non-degeneracy conditions to be found out which is beyond the reach of the usual ideal theory or *Hilbert* basis theorem. Thus, to seek for a mechanical proof of the theorem we have to seek for some other method replacing the usual ideal theory on the alike.

Such a method may be found already early works of *R. F. Ritt* on differential algebera, cf. [1, 2], For the present example the method may be described as follows.

the poly omials in the hypothesis set $\sum_1 = \{f_1, f_2, f_3\}$ have the leading variables x_4, x_8, and x_8. The first step of *Ritt* method consists in eliminating x_8 of f_3 by means of getting thus

$$f_4 = (x_5 - x_1) f_3 - x_5 f_2, \text{ or}$$

$$f_4 \equiv (x_1 x_6 + x_3 x_5 - x_1 x_3) x_7 - x_2 x_5 x_6 - x_4 x_5 (x_5 - x_1) = 0.$$

The second step of the method consists in adjioning f_4 to the original set $\sum$ to get an enlarged set $\sum_2 = \{f_1, f_2, f_3, f_4\}$ and then pick out a subset (B) called by *Ritt* a **basic set** of $\sum_2$, viz.

$$(B)\begin{cases} f_1 \equiv x_1 x_4 - x_2 x_3, \\ f_4 \equiv (x_1 x_6 + x_3 x_6 - x_1 x_3) x_7 - x_2 x_5 x_6 - x_4 x_5 (x_5 - x_1), \\ f_2 \equiv (x_5 - x_1) x_8 - x_6 (x_7 - x_2). \end{cases}$$

These enlarging and picking procedures together with some auxiliary reduction procedures may be repeated again and again if necessary but in the present case we may stop here. The basic set $\{f_1, f_4, f_2\}$ with leading variables of the polynomials x_4, x_7, x_8 is

called by *Ritt* a characteristic set of the given polynomial system, but under some less stringent conditions as originally given by *Ritt*.

The final step of *Ritt* method consists now in successive elimination form the conclusion polynomial g the variables x_8, x_7, x_4 by means of the polynomials in (B), getting thus successively:

$$(x_5 - x_1)g = x_5 f_2 + x_1 x_6 x_7 - x_2 x_5 x_6,$$

$$(x_1 x_6 + x_3 x_5 - x_1 x_3)[(x_5 - x_1)g - x_5 f_2]$$
$$= x_1 x_6 f_4 + x_5 x_6 (x_5 - x_1)(x_1 x_4 - x_2 x_3).$$

Finally,

$$(x_1 x_6 + x_3 x_5 - x_1 x_3)(x_5 - x_1)g$$
$$= x_5 (x_1 x_6 + x_3 x_5 - x_1 x_3) f_2 + x_1 x_6 f_4 + x_5 x_6 (x_5 - x_1) f_1.$$

It follows that in as much as

$$D_1 \equiv x_5 - x_1 \neq 0, \text{ and}$$
$$D_2 \equiv x_1 x_6 + x_3 x_5 - x_1 x_3 \neq 0,$$

$g = 0$ would be a consequence of $f_1 = 0$, $f_4 = 0$, and $f_2 = 0$. From the derivation of f_4 we see that under $D_1 \neq 0$, $D_2 \neq 0$, $g = 0$ would also be a consequence of the original hypothesis $f_1 = 0$, $f_2 = 0$, $f_3 = 0$. Hecen under the non-degeneracy conditions $D_1 \neq 0$, $D_2 \neq 0$ the theorem is true and the above computations constitute in fact a proof of the theorem.

It is to be emphasized that the non-degeneracy conditions are found in an automatic manner during the computations and without these $g = 0$ may not be a consequence of $f_i = 0$, $i = 1, 2, 3$ or the theorem may no more be true. Remark that $D_2 \neq 0$ means geometrically that A_1, A_3, A_5 are non-collinear. In fact, as we have already pointed out in the previous paper [3], the above part of

the Desar gues theorem will cease to be true if the triangles involved degenerate into ones with vertices collinear. If we like we may treat the degenerate cases in turn to see whether the theorem remains to be true or not. For example, let us consider the degenerate case

$$D_1 \equiv x_5 - x_1 = 0.$$

which means geometrically that A_1A_5 is parallel to the axis A_3A_4. A djoin for this $D_1 = 0$ as a new hypothesis and treat in the same manner as bove. The polynomials in the new hypothesis form now a set $\Sigma'_1 = \{f'_1, f'_2, f'_3, f'_4\}$ with

$$f'_1 \equiv x_1x_4 - x_2x_3,$$
$$f'_2 \equiv x_5 - x_1,$$
$$f'_3 \equiv (x_5 - x_1)x_8 - x_6x_7 + x_2x_8,$$
$$f'_4 \equiv x_5(x_8 - x_4) - (x_6 - x_3)x_7.$$

The first two steps consist in replacing f'_3, f'_4 by

$$f''_3 \equiv -x_6x_7 + x_2x_6,$$
$$f''_4 \equiv x_1(x_8 - x_4) - (x_6 - x_3)x_7$$

and picking out the basic set $(B)'$ given by

$$(B)' \equiv \{f'_1, f'_2, f''_3, f''_4\}.$$

It is already the characteristic set required and we readily verify that x_1x_6g is a linear combination of f'_1, f'_2, f''_3, f''_4 with polynomial coefficients.

It follows that under degeneracy condition $D_1 \equiv x_5 - x_1 = 0$ the theorem would still be true under the further non-degeneracy conditions

$$D_3 \equiv x_1 \neq 0, \quad D_4 \equiv x_6 \neq 0.$$

Which have evident geometrical meanings. If we like we may treat

in turn each of the degeneracy conditions $D_3 = 0$ and $D_4 = 0$ in the same manner as above.

The method described above is a general one and may be applied to all elementary geometrical theorems not involving order relations, though the computations will be extremely complicate for more difficult theorems than the one above. However, as the computation procedure is an algorithmic one we can write programs and practise on a computer. This has in fact already been done and difficult theorems like those of *Feuerbach* and *Morley* have been proved on the computer in this manner. Some non-trivial theorems have also been discovered by using this method which we shall not enter.

References

〔1〕 R. F. Ritt, Differential equations from the algebraic standpoint, *Amer. Math. Soc.* (1932).

〔2〕 R. F. Ritt, Differential algebra, *Amer. Math. Soc.* (1950).

〔3〕 Wu Wen-tsün, Toward mechanization of geometry—Some comments on *Hilbert's* "Grundlagen der Geometrie", *Acta Math. Scientia.* 2 (1982), 125—138.

Automation of theorem-proving*

Theorem-Proving, as a kind of mental labor, is one of the main activities of mathematies. It was originated in ancient Greek with Euclid's 《Elements》 as its representative work. However, it seems that the recognition and treatment of theorem-proving had undergone since the time of Euclid several stages of development to be explained below.

STAGE Ⅰ. Thecrem-Proving in Euclidean Fashion.

This stage of theorem-proving is characterized by the two features below:

(a) Proving is carried out in a deductive way by logical reasonings based on a Definition-Axiom-Theorem-Proof System.

(b) Each individuel theorem has an individual proof peculiar to the theorem itself.

STAGE Ⅱ. Theorem-Proving in Decartes Fashion.

The mathod of Descartes described in his 《La Geometrie》 had opened a new way of theorem-proving. The characteristic feature (a) may be turned to

(a) 'Proving of a theorem is carried out by mere computations.

---

* MM—Res. Preprints, No. 5 (1990) 5—22. Also in Workshop on Advanced Computers for Artificial Intellingence, Beijing, (1990) 11—15.

In fact, 《La Geometrie》 had shifted the main theme of theorem-proving in mathematics to the theme of problems-proving, which was in turn reduced to equations-solving. Thus, it seems that in this classic nowhere the terminology "Axiom" had been mentioned and only once the terminology "Theorem" was mentioned: "qui est un assez beau theoreme", which was deduced from computations in solving some problem about lens-making.

On the other hand the characteristic feature (b) is remaining unchanged.

STAGE Ⅲ. Thecrem-Proving in Hilbert Fashion.

In the classic 《Grundlagen der Geometrie》 Hilbert had shown how to introduce Cartesian coordinate system based on the Euclidean geometry axiomatic system via the introduction of some number system. Hilbert had thus built a bridge leading from the logical deductive method of Euclid to the computational method of Descartes. Moreover, Hilbert had changed the characteristic feature (b) of proving theorems one by one to (b)' of proving theorems class by class, viz.

(b)' A whole class of theorems is to be proved by one and the same method.

This may be seen from the fellowing theorem of the above classic:

With this notion of mechanizability we may the reformulate the Hilbert Satz 62 in the form of following

HILBERT MECHANIZATION THEOREM. The proving of the class of planar pure intersection theorems is mechanizable.

Hilbert had thus proved in the first time such a mechanization theorem. Though the theorem holds true only for a very lim-

ited class of theorems, it opened nevertheless a new era of MECHANIZATION or AUTOMATION of theorem-proving. Moreover, Hilbert had launched an ambitious program of mechanization of mathematics as a whole, in the form of consistency proving of mathematics. The program was utterly crashed owing to the appearance of the work of Goedel. However, that Hilbert program was deemed to failure is owing to the project being too ambitious rather than being unreasonable. If we restrict the class of theorems to be sufficiently smaller, then there will be higher possibility of being mechanizable. That this is really the case may be seem from the following examples.

Ex. 1. The class of theorems in arithmetic involving both addition and multiplication non-mechanizable, but the smaller class of such theorems involving addition alone is mechanizable.

The concept of mechanizability or decidability may be naturally extended to the solving of classes of problems. We have then:

Ex. 2. The solving in integers of the whole class of diophantine equations is non-mechanizable, but the same for the smaller class of linear diophantine equations is mechanizable.

Ex. 3. The word problem of the whole class of finite groups is non-mechanizable, but the same for the smaller class of free groups, or of abelian groups, or other small ones, is mechanizable.

Ex. 4. According to Tarski, the class of theorems in ordinary euclidean geometry is mechanizable, but it is seemingly in doubt for classes of theorems in other kinds of geometries.

After the above description of the different stages of theorem-

proving in past times, mainly before the appearance of computers, let us come to see the status of theorem-proving in China.

Since the year 1977 we began to work on mechanization of theorem-proving and have arrived at the follwing general mechanization theorem, cf. [WU1]:

MECHANIZATION THEOREM. The classes of theorems in various unordered geometries, euclidean or non-euclidean, ordinary, affine or projective, with fundamental elements points-lines or points-circles, etc. are all mechanizable.

We remark that by unordered geometry is meant such a one for which no axioms of order or the alike are assumed. The method of arriving at the above mechanization theorem may then be brierly described as follows. Starting from the axiomatic systems of the various geometries in question we introduce first number systems and then coordinate systems just as what Hilbert had done in his Grundlagen. A theorem in the class can then be expressed in the form of a set of polynomial equations $PS=0$ corresponding to the hypothesis and a polynomial equation $G=0$ corresponding to the conclusion. Here PS.

Satz 62. Jeder reine Schnittpunktsatz, der in einer ebenen Geometrie gilt, in der die Axiome I 1-3, II, IV* und der Pascalsche Satz gueltig sind, stellt sich durch Konstruktion geeigneter Hilfspunkte und Hilfsgeraden als eine Kombination endlich vieler Pascalscher Konfigurationen heraus.

From the statement of Satz 62 the featwe (:) geims to be remaining unchanged. However, from the passage just before the statement Eilaevt had actually illnstrated how the class of theorems involved are to be proved by mere cermuintions.

In fact, for a pare intersection theorem in the plane let the points and liner involved be intrcduces in seccession in a definite order by joining points, ferming intersections, drawing parallels, and choosing arbitrary points on a line or in the plane or arbirary line through a point or in the plane, etc. All such constructions are of a linear character. Introduce now successively parameters u_1, u_2, $\cdots$, u_d, and also bounded variables x_1, x_2, $\cdots$, x_n according to the successive constructions. Then the hypothesis of the theorem in question will be easily turned into a set of equations of the form

$$x_i = P_i / Q_i, \quad i = 1, \quad \cdots, \quad n \tag{1}$$

in which P_i, Q_i are certain polynomials in u_1, $\cdots$, u_d, and x_j with $j < i$. The conclusion will also be turned into an equation

$$G = 0, \tag{2}$$

where G is a polynomial in all u and all x. To see whether the theorem in question is true we just eliminate in succession x_n, $x_{n-1}, \cdots, x_1$, from G by means of (1), getting thus finally a fractional expression R in u alone. It is clear that the theorem will be true if and only if R is identically 0. Moreover, in the affirmative case the theorem is true only in some generic sense, or true only under some non-degeneracy conditions.

$$Q_i \neq 0, \quad i = 1, \quad \cdots, \quad n, \tag{3}$$

as seen from (1). We remark that such non-degeneracy conditions are indispensable to render (1) meaningful and should be considered as implicit hypothesis to be included in the statement of the theorem.

The above procedure of proving a pure intersection theorem is

a universal one for the whole class and is furthermore a mechanical one. Let us lay down now the following

DEFINITION. The proving of a class of theorems will be said to be MECHANIZABLE if there exists a procedure following which any theorem in the class will be seen to be true or not after a finite number of steps. By abus of language we shall say also that the class of theorems is mechanizable.

We remark that here we have used the terminology MECHANIZABLE rather than DECIDABLE usually used in mathematical logic. It seems that it was Hao Wang, reknown for his pioneering contributions in mechanical theorem-proving, who first introduced the notion of MATHEMATICS-MECHANIZATION and we have followed this usage. s a polyncmial set and G a polynomial all in certain variables $x_1, \cdots, x_n$ with ceefficionts in certain field K intrinsic to the geometry in question. We remark that owing to the inorderedness of the geometry only equations and inequations will be involved but not at all inequalities, and the intrinsic field K is also an unordered one unlike the ordered field n the case considered by Tarski.

In such a formulation a ZERO of the polynomial set PS corresponding to the hypothesis or a solution of $PS=0$ is then nothing else but a geometric configuration verifying the hypothesis, eventually in some imaginative extended space. Denote the totality of zeres of PS by $Zero\ (PS)$. Then to prove the theorem is equivalent to the determination of the ubset of $Zero\ (PS)$ on which the conclusion polynomial G should be equal to 0. We are thus led to the problem of solving the system of equations $PS=0$ or one of determining the structure of $Zero\ (PS)$. This latter

problem has been solved in the form of some decomposition theorems which express *Zero* (*PS*) as union of components $Zero_i$. Each of such $Zero_i$ is so well-behaved that the zeros of this set can be readily determined if equired. As a consequence it is easy to verify whether $G=0$ on this component $Zero_i$ or not by mere computations which is a direct generalization of the procedure of Hilbert in his mechanization theorem. For more details we refer to the relavant papers of the auther.

It is seen thus the theorem-proving, at least in the case covered by the above general mechanization theorem, may be considered as a mere application of our method of equations-solving. This may be true in other occasions too. In fact, the above mechanization theorem has been extended to some classes of theorems in various kinds of differential geometries and the theorem-proving for such classes is reduced to the equations-solving of systems of differential equations, or the determining of the structure of zero-set of the corresponding sets of differential polynomials.

The mathematics in our ancient China was of a highly mechanical character. Most of the results were expressed in a form equivalent to now-a-days algorithms which may be readily turned into programs to be carried out on a computer. Moreover, the main concern of our ancestors was problem-solving rather than theorem-proving. The problem-solving was then treated by reducing it to equations-solving, quite in the same spirit of Descartes. In fact, our method of solving arbitrary polynomial equations $PS=0$ in the form of determining the structure of *Zero* (*PS*) may be considered just as a systematization and modernization of the method developed around 13-th century in ancient Chi-

na, in adopting some modern concepts and techniques due to J. F. Ritt in 1950. Our treatment of theorem-proving follows thus closely the line of development

Ancient China ⟶ *Descartes* ⟶ *Hilbert*

In the spirit of this line of development we have also laid down the following program as our future goal of theorem-proving and problem-solving:

MATHEMATICS-MECHANIZATION PROGRAM. Cover as much as possible the whole of mathematics by small domains each of which is mechanizable.

REFERENCES

[WU1] Wu Wen-tsun, Basic Principles of Mechanical Theorem Proving in Geometries (Part on Elementary Geometries), (in Chinese), Science Press, (1984).

A survey of developments of mathematics mechanization in China*

Abstract

The present survey is devoted to activities in China on theorem-proving and equations-solving, the two main topics of MM (Mathematics Mechanization), restricted to works mainly due to the MM group in the Institute of Systems Science, Academia Sinica. The contributions on equations-solving in ancient China are also briefly described.

1 Mathematics Mechanization in Ancient Times

There are two main kinds of activities in mathematics, viz. equations-solving and theorem-proving. The latter one was typically of Greek origin and Euclid's Elements served as its representative work. In this classic a deductive system was established and the exposition bore the style of the form Definition-Axiom-Theorem-Proof. On the other hand the equations-solving was o-

* Chinese Mathematics into 21st century, (Eds. W. T. Wu & M. D. Cheng) Peking University Press, Beijing, (1991) 15—40.

riginated and well-developed in ancient China. In fact, the ancient Chinese were mainly devoted to solve concrete problems arising in practice which led naturally to problems of equations-solving. The representative works were the classic "Nine Chapters in Arithmetic" and its "Annotations" by Liu-Hwei, which, in accordance with the main spirit, adopted an exposition bearing the style of the form Example-Answer-Method & Result-Explanation. The item Method & Result was usually expressed in a form which now-a-days we would like to call it an **ALGORITHM.** The item Explanation usually gave the underlying reasons of the algorithm somewhat like a now-a-days correctness-proof. The sources of such equations were tow-fold. One of such sources was rudimentary commerce or goods-exchange which led to the Excess-Deficiency Algorithm in Chapter **7** of "Nine Chapters" corresponding to, in modern language, the method of solving a system of two linear equations in two unknowns. It was developed into a general method of solving simultaneous linear equations which was the main concern of Chapter **8** of "Nine Chapters". This method was the same as the one bearing the name of Gaussian Elimination Method in modern writings, which, what seems to be legitimate, would be more accurately to be called Nine-Chapters-Elimination-Method. A second source of equations-solving was from problems of measurements. The simplest one of such problems consists in extracting square roots for the determination of the hypothenus (e. g. the sun-observer distance) from the known values of the two sides of a right-angled triangle (e. g. the sun-height and level-distance of sun-observer). It requires the extraction of square roots which is equivalent to the solving of a quadrat-

ic equation of the form $x^2 = A$. It is perhaps owing to this origin that throughout the incessant development of equations-solving up to the method of numerical solution of any higher degree algebraic equations during the Song dynasty the Chinese ancestors called equations-solving simply roots-extraction (with extra terms). The discovery of the notion of Tian-Yuan or Heavens-Element equivalent to the present-day notion of Unknown during the Song and Yuan dynasties furnished a systematic method of establishing equations for given problems. The two lines of developments of equations-solving were also rendered possible to merge into one which was closer to algebra in modern sense. It cultimated in the classic of Zhu Szejze bearing the title "The Jade Mirror of Four Elements" which gave an elimination method of solving systems of algebraic higher degree equations up to four unknowns. The following, expressed in modern languages, was one of the representative examples given in Zhu's classic:

Problem. For a right-angled triangle the hypothenus z and the two sides x and y are connected by the relations

$$x \cdot y \cdot z - x \cdot y^2 - z - x - y = 0, \tag{1.1}$$

$$x \cdot z - x^2 - z - y + x = 0, \tag{1.2}$$

$$x^2 + y^2 = z^2. \tag{1.3}$$

Find x, y, and z.

Method of Zhu(a little simplified).

Eliminate y from(1.1)by means of (1.2), (1.3)and remove the factor x one gets

$$-2 - z - 2 \cdot z^2 + x + x \cdot z + x \cdot z^2 + x^2 - x^2 \cdot z = 0. \tag{1.4}$$

Eliminate y from (1.2)by means of (1.3)and remove the factor x one gets

$$-2z-2\cdot z^2+2\cdot x+4\cdot x\cdot z+x\cdot z^2$$
$$-2\cdot x^2-x^2\cdot z+x^3=0. \quad (1.5)$$

Write(1.4)and (1.5)as equations in x and eliminate x^3, $x^2$from (1.5) by means of (1.4), one gets successively equations in x^2 and x:

$$(-3-z+z^2)\cdot x^2+(4+3\cdot z-z^2-z^3)\cdot x$$
$$-2\cdot z+2\cdot z^3=0, \quad (1.6)$$
$$(7+3\cdot z-z^2)\cdot x-(6+7\cdot z+3\cdot z^2-z^3)=0$$

Eliminate finally x from(1.6)and(1.4)one gets then an equation in z alone:

$$z^4-6\cdot z^3+4\cdot z^2+6\cdot z-5=0. \quad (1.7)$$

Solving, one gets one root $z=5$. Substituting in (1.6)one finds $x=3$. Substituting the values of x and z in (1.2), one finds $y=4$. A solution of the problem is thus given by$(x, y, z)=(3,4,5)$.

The limitation of Zhu's method to utmost 4 unknowns was due to the fact that the Chinese ancients made all calculations on counting boards. Otherwise the method is entirely general and may be considered as a direct generalization of Nine-Chapters-Elimination-Method. Moreover, all such methods are highly mechanical in character and may be turned readily into computerized programs if one likes. This is indeed in high contrast with the theorem-proving dated from the ancient Greece which is exceedingly non-mechanical in character.

A turning point occurred in 17th century when Descartes, and also Fermat, created the method of analytic geometry. It seems that Descartes, as shown by his monumental classic La Géomètrie", paid great emphasis on solving of concrete problems, which led naturally to equations-solving, while theorem-proving

was somewhat neglected. In this respect the attitude of Descartes seems to be much in the same spirit of our Chinese ancestors. In any way the creation of analytic-geometrical method opened the way of mechanization, not only of equations-solving, but also theorem-proving, no matter how late such mechanization does begin to occur.

2 Discovery of Mechanization Methods for Theorem-Proving

The mathematics mechanization in a strict sense may be considered to begin since the appearance in **1899** of Hilbert's classic "Grunlagen der Geometrie". Hilbert was the doubtless founder of the modern axiomatization of mathematics. However, it was first time pointed out by Wu([Wu**3**, **5**, **6**]) that Hilbert may also be considered as the pioneer of the actual mechanization of mathematics, especially in the realm of theorem-proving. In fact, in the end of the main body of the above classic, i. e. in the end of Chap **6**, there was an obscure passage which was turned into the form of a theorem in the later editions, viz.

Theorem 62. Each pure intersection-point theorem in a Pascalian geometry, if true, can always be proved by the aid of suitably constructing auxiliary points and lines, as consequence of a combination of a finite number of Pascal configurations.

It seems that this theorem has never been taken into notice, in spite of the fact that Hilbert's classic was widely and seriously studied for all the time. It was Wu who first recognized the significance of this theorem and reformulated it in the following form:

Hilbert Mechanization Theorem. In a planar Pascalian geometry there is a mechanical procedure which permits to prove or disprove in a finite number of steps any pure intersection-point theorem of constructive type under certain subsidiary non-degeneracy conditions also generated mechanically during the procedure.

It would be better to elucidate the actual meaning of this theorem by a concrete example. Let us consider thus the following

Theorem of Gauss(Werke, Bd. 4, S. 385—392). For a complete quadrilateral consisting of 4 lines $L_1, \cdots, L_4$ intersecting in 6 points L_{ij}, the mid-points M_1, M_2, M_3 of the 3 diagonals L_{12} L_{34}, $L_{13}L_{24}$, and $L_{14}L_{23}$ are co-line.

The occurrence of the concept mid-point would surpass the range of applicability of the above mechanization theorem which does not matter, however. The mechanical procedure of Hilbert still works and runs as follows.

Take an arbitrary point L_{12} as the origin O. Take arbitrary lines L_1 and L_2 through L_{12} as the x-and the y-axis. Take on the x-axis L_1 arbitrary points L_{13} and L_{14} to be given the coordinates $(U_1, 0)$and $(U_2, 0)$with U_1 and U_2 as parameters. Take on the y-axis L_2 arbitrary points L_{23} and L_{24} to be given the coordinates $(0, U_3)$and $(0, U_4)$. Draw now the line joining L_{13}, L_{23} and also the line L_4 joining L_{14} and L_{24}. Form the intersection point L_{34} of lines L_3 and L_4 and give it the coordinates (Y_1, Y_2) for which Y_1, Y_2 are no more arbirtary parameters but bounded with coordinates U of points already taken. The equations determining Y_1, Y_2 may be readily put in the following form:

$$Y_1 = U_1 \cdot U_2 \cdot (U_4 - U_3) / (U_1 \cdot U_4 - U_2 \cdot U_3), \tag{2.1}$$

$$Y_2 = U_3 \cdot (U_1 - Y_1) / U_1. \tag{2.2}$$

Draw now the line $L_{12}L_{34}$ and take the mid-point M_1 to be given the coordinates(Y_3, Y_4) so that

$$Y_3 = Y_1/2, \tag{2.3}$$

$$Y_4 = Y_2/2. \tag{2.4}$$

Draw next the line $L_{13}L_{24}$, take the mid-point M_2, and give it the coordinates(Y_5, Y_6)with

$$Y_5 = U_1/2, \tag{2.5}$$

$$Y_6 = U_4/2. \tag{2.6}$$

Finally draw the line $L_{14}L_{23}$ with mid-point $M_3(Y_7, Y_8)$ such that

$$Y_7 = U_2/2, \tag{2.7}$$

$$Y_8 = U_3/2. \tag{2.8}$$

The conclusion of the theorem is equivalent to $G = 0$, with the conclusion polynomial (abbr. henceforth as pol) Ggiven by

$$G = Y_3 \cdot Y_6 + Y_5 \cdot Y_8 + Y_4 \cdot Y_7 - Y_3 \cdot Y_8 - Y_4 \cdot Y_5.$$

The proof of theorem now runs as follows.

Eliminate Y_8 in $G_8 = G$ by means of(2.8) to get a pol G_7 in U and $Y_1, \cdots, Y_7$ alone. Eliminate next Y_7 in G_7 by means of (2.7)to get a pol G_6 in U and $Y_1, \cdots, Y_6$ alone, etc. Finally, after eliminating Y_1 from G_1 by means of(2.1) one would get a pol G_0 in U alone which turns out to be identically 0. This proves the theorem. Moreover, it shows that the theorem is proved to be true only under the conditions $U_1 \neq 0$ and $U_1 \cdot U_4 - U_2 \cdot U_3 \neq 0$ which are the non-degeneracy ones as stated in the Mechanization Theorem and are found automatically with U_1, $U_1 \cdot U_4 - U_2 \cdot U_3$ as the denominators occurring in the equations(2.1) – (2.8).

The method of proof is a general one applicable to the whole class of theorems possessing the following properties:

Coordinates can be so chosen such that parameters U_1, U_2, $\cdots$, and bounded variables Y_1, Y_2, $\cdots$, will be introduced one by one with hypothesis of the theorem expressible in the form

$$Y_i = P_i / Q_i,\ i = 1, 2, \cdots,$$

the P_i and Q_i being pols in U and Y_j alone for $j < i$. The conclusion is also expressible in the form $G = 0$ with the conclusion pol G one in U and Y. Eliminate now the Y's in G successively in the reverse order by means of these hypothesis equations. The theorem in question will then be seen to be true or un-true according to whether the final result is identically 0 or not. In the affirmative case the theorem is true GENERICALLY under the NON-DEGENERACY CONDITIONS $Q_i \neq 0$.

The method is seemingly a very trivial one which permits, however, to prove already quite non-trivial theorems. In year 1977 Wu discovered, being entirely ignorant of the above Hilbert's result at that time, a more general method of much wider applicability for mechanical geometry theorem-proving as follows (cf. [Wu, 1, 2, 3])

Let K be a field of characteristic 0. For any pol in variables U_1, $\cdots$, U_d and Y_1, $\cdots$, Y_e on coefficients in K let c be the greatest subscript of Y such that Y_c actually occurs in G. One can then write G in the form

$$G = I \cdot Y_c^D + \text{lower degree terms in } Y_c,$$

in which all coefficients of Y_c-powers, if not 0, will be pols in U and Y_i with $i < c$ alone. We call then I the INITIAL of G and Y_c the **LEADING VARIABLE of** G. Suppose now a theorem

T be such that with coordinates suitably chosen, the hypothesis are expressible as $H_i = 0$, $i = 1, \cdots, e$ with H_i pols having Y_i as leading variables:

$$H_i = I_i \cdot Y_i \cdot D_i + \text{lower degree terms in } Y_i, \qquad (2.\ 9)$$

If in particular H_1 is irreducible as a pol in Y_1 on the field $K_0 = K(U_1, \cdots, U_d)$, H_2 is irreducible as a pol in Y_2 on the field $K_1 = K_0(Y_1)$ extended by adjoining Y_1 to K_0 by means of $H_1 = 0$, etc, Then the hypothesis set forming of $H_1, \cdots, H_e$ will be said to be IRREDUCIBLE. Suppose that $G = 0$ with$G$ a pol in $U$and Yon K be the conclusion of the theorem T. To see whether T is true of not let us just divide G successively by H_l, $\cdots H_2$, H_1 to remove the Y's as much as possible. We get then, for some non-negative integers S_i, and also pols A_i, a formula of the form

$$I_1^{S_1} \cdots\cdots I_e^{S_e} \cdot G = A_1 \cdot H_1 + \cdots + A_e \cdot H_e + R, \quad (2.\ 10)$$

in which R is a pol in U and $Y$with degree in each Y_i less than the corresponding degree D_i in H_i. We call R the REMAINDER of the pol G w. r. t. the hypothesis-set $HYP = \{H_1, \cdots, H_e\}$ and write it as

$$R = \text{Remdr}\ (G/HYP\}.$$

From the remainder formula (2. 10) we have the following

Mechanical Theorem-Proving Principle (abbr. MTP). If the remainder R off G w. r. t. the hypothesis-set HYP is identically, 0, then the theorem T in question is GENERICALLY true under the non degeneracy conditions $I_i \neq 0$, $i = 1, \cdots, e$. Moreover, the condition $R = 0$ is not only sufficient but also necessary for the theorem T to be generically true under the above non-degeneracy conditions in case the hypothesisset is irreducible.

The above principle serves as the basis for proving, or even disproving, theorems in elementary geometries. We remark that theorems in elementary geometries can be usually, after some simple manipulations, put in the form of (2. 10). Even trigonometrical identities can be transformed into the form of (2. 10) by treating the various trigonometric functions as separate variables connected by algebraic relations. We refer to the examples given in the earliest paper [Wul] of Wu on this subject which already show clearly how efficient the method could be. The method may even be served to discover new non-trivial ones since once one guesses some possible geometrical truth, one may just make verifications by the above principle to see whether it is really true or not. Among the earliest new theorems discovered in this way we may mention the Pascal Conic Theorem about chord intersections of 6 points on a conic, cf. e. g. [Wu2, 3]. The following example is a further illustration of this principle ([Wu9]).

Example. Consider a plane moving over a fixed one and occupy 4 different positions. To move from i-th position to the j-th position there is associated a center of rotation of POLE P_{ij} and an angle of rotation R_{ij} mod $2 \cdot p_i$. Suppose that all such poles P_{ij} are situated in the finite part of the fixed plane so that we have a 6-pole-configuration. In some particular cases the configuration may degenerate into a complete quadrilateral with 3 pairs of diagonals $P_{12} P_{34}$, $P_{13} P_{24}$, and $P_{14} P_{23}$. In view of the theorem of Gauss we may ask ourselves whether this theorem remains true for the general 6-pole-configuration.

To see this let us first remark that the configuration will be

completely determined by the positions of P_{12}, P_{13}, P_{14} and the 3 rotation angles R_{12}, R_{13}, and R_{14}. Let us take thus coordinates such that

$$P_{12} = (0, 0), \quad P_{13} = (U_1, 0), \quad P_{14} = (U_2, U_3),$$

and set

$$\text{Tan } (R_{12}/2) = U_5 : U_4, \quad \text{Tan } (T_{13}/2) = U_7 : U_6,$$

$$\text{Tan } (R_{14}/2) = U_9 : U_8.$$

From fundamental theorems in planar kinematics we know that

$$\angle (P_{12}P_{13}, P_{12}P_{23}) = \angle (P_{12}P_{14}, P_{12}P_{24})$$
$$= R_{12}/2 \bmod p_i, \text{ etc.}$$

Let the coordinates of P_{23}, P_{24}, P_{34} and the mid-points M_1, M_2, M_3 of the diagonals $P_{12}P_{34}$, $P_{13}P_{24}$, and $P_{14}P_{23}$ be given by

$$P_{23} = (Y_1, Y_2), \quad P_{24} = (Y_3, Y_4), \quad P_{34} = (Y_5, Y_6),$$
$$M_1 = (Y_7, Y_8), \quad M_2 = (Y_9, Y_{10}), \quad M_3 = (Y_{11}, Y_{12}),$$

The geometrical conditions will then be turned into a hypothesis set consisting of 12 equations which, after simple manipulations, will be of the form $H_i = 0$, $i = 1, \cdots, 12$, with

$$H_1 = + Y_1 \cdot U_7 \cdot U_4 - Y_1 \cdot U_6 \cdot U_5 - U_7 \cdot U_4 \cdot U_1,$$
$$H_2 = - Y_2 \cdot U_7 \cdot U_4 + Y_2 \cdot U_6 \cdot U_5 + U_7 \cdot U_5 \cdot U_1,$$
$$H_3 = - Y_3 \cdot U_9 \cdot U_4 + Y_3 \cdot U_8 \cdot U_5 - U_9 \cdot U_5 \cdot U_3 + U_9 \cdot U_4 \cdot U_2,$$
$$H_4 = - Y_4 \cdot U_9 \cdot U_4 + Y_4 \cdot U_8 \cdot U_5 + U_9 \cdot U_5 \cdot U_2 + U_9 \cdot U_4 \cdot U_3,$$
$$H_5 = - Y_5 \cdot U_9 \cdot U_6 + Y_5 \cdot U_8 \cdot U_7 - U_9 \cdot U_7 \cdot U_3 + U_9 \cdot U_6 \cdot U_2 \cdots - U_8 \cdot U_7 \cdot U_2 + U_8 \cdot U_7 \cdot U_2 - U_8 \cdot U_7 \cdot U_1,$$
$$H_6 = - Y_6 \cdot U_9 \cdot U_6 + Y_6 \cdot U_8 \cdot U_7 \cdots + U_9 \cdot U_7 \cdot U_2 - U_9 \cdot U_7 \cdot U_1 + U_9 \cdot U_6 \cdot U_3,$$

$$H_7 \quad = 2 \cdot Y_7 - Y_1,$$

$$\cdots\cdots\cdots\cdots$$

$$H_{12} \quad = 2 \cdot Y_{12} - Y_2 + U_3.$$

The conclusion to be verified is $G = 0$ with the conclusion pol given by

$$G = Y_7 \cdot Y_{10} + Y_9 \cdot Y_{12} + Y_8 \cdot Y_{11} - Y_{10} \cdot Y_{11} - Y_7 \cdot Y_{12} - Y_8 \cdot Y_9.$$

It is readily verified , HYP being the polset $\{H_1\}$, that

$$\text{Remdr } (G/HYP) = 0.$$

We have thus discovered a new theorem generalizing the theorem of Gauss. The non-degeneracy conditions are given by $I_i \neq 0$ with I_i the initial of H_i. It is readily seen that $I_i = 0$ corresponds to the degeneracy of the 6-pole-configuration into a complete quadrilateral and the theorem is already known to be true in this case owing to above theorem of Gauss.

In general, if we want to know whether a theorem remains to be true in some degenerate case $I = 0$, we may just add I to the original hyp-set to form a new one HYP' and try to see whether the conc-pol is still 0 at that time. Now it may happen, as is usually the case, that the hyp-set HYP' will be no more in the well-ordered form like (2. 9). Some preliminary manipulations are then necessary to bring HYP' again to a well-ordered form. This and other considerations induce us to seek a more general setting of treatments as we shall see in the next section.

3 A General Method of Mechanization for Equations-Solving

Now there are theorems in elementary geometries for which it is rather complicate to bring the hypothesis set in some coordinate system to the form of (2. 9). There are also occasions to consider problems dealing with sets of pols entirely arbitrary. It is also important to extend the whole theory to the case of differential geometries. Thus, after the completion of the work [Wul], Wu saw the importance of the works of Ritt, at that time somewhat lying in the obscurity in the mathematics world, and began to study seriously the second book of Ritt, [R2]. The study was a fruitful one and brought about fertile results in diverse directions which will be explained below.

Let us consider a set of variables $X_1, X_2, \cdots, X_m$ and a field K of characteristic 0 fixed henceforth. Any polynomial in the ring

$$K\ [X_1, \cdots, X_m]$$

will be simply called a POL. A finite set of non-zero pols is then called a POLSET. Any point $A = (A_1, \cdots, A_m)$ with A_i in some extension field K' of K is called a K'-ZERO or simple a ZERO of a pool P if P becomes 0 when X_i in P are replaced by A_i. The set of K'-zeros of all pols in a polset PS will be denoted by K'-Zero (PS) or simply Zero (PS) if K' is taken to be arbitrary or unnecessary to be specified . If G is any other pol, then we shall also set

$$\text{Zero}\,(PS/G) = \text{Zero}(PS)/\text{Zero}(G).$$

We are now interested in solving the system of equations $P_i = 0$ with P_i running over the pols in a given polset PS, and in the study of the structure of the set Zero (PS). For this purpose we shall introduce the notions originated in Ritt's works as follows.

For any pol P let us set

c = greatest subscript of X for which
X_c actually appears in P,

d = highest degree of X_c in P.

For a non-zero constant pol we set simply $c=0$ and $d=0$. We call c the CLASS and d the DEGREE of P to be denoted by Cls (P) and Deg (P) resp. A non-zero pol Q will be said to be REDUCED w.r.t P if P is not a constant one and, for $c=$ Cls $(P) > 0$, the degree of Q in X_c is less than Deg (P). We introduce now among the non-zero pols a partial ordering $<$ such that $P < Q$ if either Cls $(P) <$ Cls (Q) or Cls $(P) =$ Cls (Q) and Deg $(P) <$ Deg (Q). If neither is $P<Q$ nor is $Q<P$ in the above ordering then we say that P, Q are INCOMPARABLE in ORDER and we write then $P <=> Q$.

Consider now a finite sequence of non-zero pols

$$(P) \qquad P_1, P_2, \cdots, P_r.$$

We call (P) an ASCENDING SET if one of the following cases holds true:

(1) $r=1$ and Cls $(P_1) = 0$ or P_1 is a non-zero constant in K. The asc-set in this case is then said to be TRIVIAL of CONTRADICTORY.

(2) $0 <$ Cls $(P_1) < \cdots <$ Cls (P_r)

and each P_i is reduced w.r.t. any preceding P_j with $j<i$.

We say that a pol G *is REDUCED w. r. t. an asc-set* (P) if G is reduced w. r. t. any pol in (P).

Let (P) be now a non-trivial asc-set as above and G any pol, then by successively dividing by pols of G in the reverse order, we get a formula of the form

$$J \cdot G = \sum_i A_i \cdot P_i + R,$$

in which J is a power-product of initials I_i of P_i, and A_i, R are pols with R reduced w. r. t. (P). We can choose exponents of I_i in J to be as small as possible. The pol R is then uniquely determined and will be called the REMAINDER of G w. r. t. (P), to be denoted by

$$R = \text{Remdr}\,(G/(P)).$$

We now introduce a partial ordering $<$ among all the asc-sets in the following way. A trivial asc-set will be always $<$ a non-trivial one . Let us consider, besides the non-trivial asc-set (P) as given above, a second non-trivial asc-set

$$(Q) \qquad Q_1,\ Q_2,\ \cdots,\ Q_s.$$

Then we say P is $< Q$ if one of the following conditions holds true:

(1)′ There is some k such that $P_i \Leftrightarrow Q_i$ for $i < k$, while $P_k < Q_k$.

(2)′ $r > s$ and $P_i \Leftrightarrow Q_i$ for $i = 1, \cdots, s$.

If for two asc-sets (P) and (Q) neither is $(P) < (Q)$ nor is $(Q) < (P)$, then we say that (P) and (Q) are INCOMPARABLE in ORDER and we write then $(P) \Leftrightarrow (Q)$.

Consider now any polset PS. Any asc-set of lowest order contained in PS will be called a BASIC-SET (abbr. bas-set) of PS. All such basset of PS are clearly incomparable in order each

other. It is therefore legitimate to define for any two polsets PS_1 and PS_2 whether PS_1 is $< PS_2$ or PS_2 is $< PS_1$ according to their bas-sets.

For any polset PS we form now a certain particular asc-set CS according to the following scheme:

$$\begin{aligned} PS = \quad & PS_0 \subset PS_1 \subset \cdots \subset PS_s \\ & BS_0 > BS_1 > \cdots > BS_s = CS \\ & RS_0 \quad RS_1 \quad \cdots \quad RS_s \quad = \text{Empty} \end{aligned} \tag{S}$$

The above scheme (S) is formed in the following way:

Each BS_i is a bas-set of PS_i.

Each RS_i is the polset consisting of all non-zero remainders of pols in $PS_i \setminus BS_i$ w. r. t. BS_i.

Each polset PS_i is the union of PS_j and RS_j where $j = i - 1$.

$$PS_i = PS_j + RS_j, \quad j = i - 1.$$

We have then

$$BS_0 > BS_1 > \cdots$$

a sequence steadily decreasing in order and the construction should end in a finite number of steps so that in certain stage s we should have

$$RS_s = \text{Empty}.$$

DEF. The corresponding bas-set $BS_s = CS$ in the above scheme (S) is called a CHARACTERISTIC SET (abbr. char-set) of the given polset PS.

Well-Ordering Principle. Let I_i be the initials of pols in CS and J be the product of all these I_i. Then:

$$\text{Zero}\ (CS/J) \subset \text{Zero}\ (PS) \subset \text{Zero}\ (CS),$$

or more precisely,

$$\text{Zero}(PS) = \text{Zero}(CS/J) + \sum_i \text{Zero}(PS_i)). \qquad (\text{I})$$

In the formula (I) each PS_i is the enlarged polset obtained from PS by adjoining to it the initial I_i of the i-th pol of CS.

Now each polset PS_i in (I) is of lower order than PS. Hence if we apply (I) to each of PS_i and proceed in the same way again and again we should stop in a certain stage and arrive at the following.

Zero Decomposition Theorem (Weak Form) There is an algorithmic procedure which permits to give for any polset PS a decomposition of the following form:

$$\text{Zero}(PS) = \sum_k \text{Zero}(CS_k/J_k). \qquad (\text{II})$$

In the formula each CS_k is an asc-set and J_k is the product of all initials of pols in CS_k.

Consider now a non-trivial asc-set(P)as before. we shall say that (P) is REDUCIBLE at a certain stage i if we have a relation of the form

$$H_i \cdot P_i = P_i' \cdot P_i'' + \sum_j A_j \cdot P_j,$$

in which P_i', P_i'' are pols having same class as P_i while H_i is a pol with smaller class and is reduced w.r.t. the partial asc-set ($k = i - 1$)

$$(PP_k) \qquad P_1, \cdots, P_k.$$

The $\sum_j$ means summation over j from 1 to k with A_j some pols.

If(P) is not reducible at any stage, then(P)is said to be IRREDUCIBLE. In that case we can extend the field K to a field IK by successively adjoining $Z_1, Z_2, \cdots, Z_m$ to K with each Z_c either a transcendental element or an algebraic element defined by the equation $P_i = 0$, when c is the class of some pol P_i in (P).

The point $Z=(Z_1, Z_2, \cdots, Z_m)$ in the space IK^m is then called a GENERIC POINT of the irreducible asc-set (P). We have now the following.

Theorem. W.r.t. an irreducible asc-set (P) with a generic zero Z we have for any pol G:

$$\text{Remdr}(G/(P)) = <===> Z \in \text{Zero}(G).$$

A point Z' in K'^m where K' is any extension field of K is called a SPECIALIZATION of Z if for any pol P,

$$Z \in \text{Zero}(P) ==\Rightarrow Z' \in \text{Zero}(P).$$

The set of all specializaitons of Z forms then an algebraic variety with Z as generic point in the usual sense. This variety is completely determined by Z, which, if it is a generic point of an irreducible asc-set (P), is in turn completely determined by (P). We shall call accordingly this variety as one ASSOCIATED TO(P), to be denoted by Var$((P))$.

Consider now an asc-set (P) which is reducible at certain stage i as before. Then we have a decomposition of the form

$$\text{Zero}((P)) = \text{Zero}((P)'/H_i) + \text{Zero}((P)''/H_i) + \text{Zero}(((P+)),$$

in which $(P)'$, $(P)''$ are the sequences obtained from (P) in replacing P_i of (P) by P' and P'' resp., while $(P+)$ is the polset enlarging (P) by adjoining to it the pol H_i. Applying now this procedure to the decomposition (II) whenever some asc-set CS_k is reducible and proceed in the same manner further and further as long as possible, we shall arrive at the following theorems:

Zero Decomposition Theorem(Strong Form). There is an algorithmic procedure which permits to give for any d-polset PS a decomposition of the form

$$\text{Zero}(PS) = \sum_k \text{Zero}(IRR_k/J_k), \qquad \text{(III)}$$

in which each IRR_k is an irreducible asc-set and each J_k is the product of all initials of pols in IRR_k.

Variety Decomposition Theorem. For any polset PS we have a decomposition of the form

$$\mathrm{Zero}(PS) = \sum_k \mathrm{Var}(IRR_k), \qquad (\mathrm{IV})$$

in which $\mathrm{Var}(IRR_k)$ is the algebraic variety associated to certain irreducible asc-set IRR_k. Moreover, this decomposition can be rendered uncontractible, and hence also unique, by mere computations.

The above decomposition theorems and the formulas (I)—(IV) give a complete overview of the structure of the zero-set of any polset. They form also the basis of a constructive theory of algebraic varieties. Moreover, since each asc-set possesses a form which permits to solve the corresponding equations one by one, they furnish us also a general method of solving arbitrary systems of algebraic equations which may be considered as a direct and more systematic generalization of the elimination method created by Zhu Szejze and his predecessors. Cf. [Wu9 – 11, 16, 20]. We remark however that there are various modified improved forms of the schems (s)for the formation of char-set, cf. e.g. [Wu16, 21]. We remark also that Gao in a paper [Gao4] has introduced the notion of Minimal Char-Set which is uniquely determined by the given polset.

The above theory has also been extended by Wu to the case of differential polsets, along the lines of development by M. Ch. Riquier, M. M. Janet, E. Cartan, J. M. Thomas, and particularly J. F. Ritt. The results are analogous to those in the above polynomial case but, of course, much more complicated in that integrability

conditions or passivity in the sense of Riquier should be taken into consideration. We shall not enter into it here but refer to the relevant papers of Wu, particularly the paper[Wu20] in MM-Preprints No. 3.

4 Basic Principles of Mechanization of Theorem Proving

The various decomposition theorems given in Sect 3 will naturally give rise to mechanization methods for the proving of theorems which are more general than the one in Sect 2 and are complete in a certain sense. Let us restrict ourselves first to the case of elementary geometries, i.e. geometries for which no notions of differentiations are involved. On the other hand it is immaterial whether the geometry concerned is the ordinary, the affine, the projective, or any one else. To make precise let us define a THEOREM in such geometries as a couple consisting of a polset called HYPOTHESIS SET (abbr. hyp-set) and a pol called CONCLUSION POL (abbr. conc-pol). We lay down now the following

DEF. Let the hyp-set and the conc-pol of theorem T be resp. HYP and $CONC$. Then we say:

(1) T is TRUE if

$$\mathrm{Zero}(HYP) \subset \mathrm{Zero}(CONC).$$

(2) T is GENERICALLY TRUE under NON-DEGENERACY CONDITIONs

$$DEG_i \neq 0$$

for DEGENERACY POLs (abbr. deg-pol) DEG_i if

$$\mathrm{Zero}(HYP/\prod DEG_i) \subset \mathrm{Zero}(CONC).$$

(3) T is TRUE on a part PV of the algebraic variety Zero (HYP) if

$$PV \subset \text{Zero}(CONC).$$

We remark that a zero in Zero(HYP) is nothing but a geometrical configuration verifying the hypothesis of then theorem T and Zero(HYP) is just the algebraic variety of all such geometrical configurations.

For the hyp-set HYP of a theorem T let us form now a char-set CS according to scheme (S) in Sect 3. Let the initials of pols in CS be I_i and J their product. Let $HY\,P_i$ be the enlarged polsets obtained frome $HY\,P$ by adjoining to it I_i. Then by the well-ordering Principle we have:

$$\text{Zero}(HYP) = \text{Zero}(CS/J) + \sum_i \text{Zero}(HYP_i).$$

From now the remainder R of the conc-pol $CONC$ of theorem T w.r.t. CS:

$$R = \text{Remdr}(CONC/CS).$$

Suppose that $R = 0$ as a pol. Then we see from the remainder formula in Sect 3 that $HYP = 0$ would imply that $CONC = 0$ so far $J \neq 0$ or no I_i is 0, i.e.

$$\text{Zero}(HYP/J) \subset \text{Zero}(CONC).$$

The theorem T is thus seen to be generically true under the non-degeneracy conditions

$$I_i \neq 0.$$

If the asc-set CS is furthermore irreducible, then the generic zero Z of CS is a zero of CS for which no I_i is 0 and so it is also a zero of HYP.

If the theorem T is generically true under the nondegeneracy conditions $I_i \neq 0$ so that in particular Z is a zero of $CONC$ by the for-

mula above, then it follows from the theorem about generic zero in Sect 3 that we have necessarily $R=0$ as a pol. Hence we have:

Principle of Mechanical Theorem Proving or MTp-Principle I(Weak Form). If the remainder R of $CONC$ w.r.t. a char-set CS of HYP is 0, then the theorem T with hyp-set HYP and conc-pol $CONC$ is generically true under the non-degeneracy conditions $I_i \neq 0$ for which I_i are the initials of pols in CS. Moreover, if the char-set CS is irreducible, then the above condition $R=0$ is also a necessary one for then theorem T to be generically true under the above non-degeneracy conditions.

Let us now apply the zero-decomposition theorem (weak form)(Ⅱ) to the hyp-set HYP so that

$$\text{Zero}(HYP) = \sum_k \text{Zero}(CS_k/J_k),$$

in which each CS_k is an asc-set and J_k is the product of initials of pols in CS_k. Let R_k be the remainder of $CONC$ w.r.t. CS_k. Then we have the following

MTP-PRINCIPLE I′(General Weak Form). if $R_k=0$, then the theorem T is true on the part Zero(CS_k/j_k) of the whole variety Zero(HYP) of geometrical configurations verifying the hyp-set HYP. The condition $R_k=0$ is also necessary if CS_k is irreducible.

More generally, let us apply the variety-decomposition theorem to $HY\ P$ so that we have

$$\text{Zero}(HYP) = \sum_k \text{Var}(IRR_k),$$

in which IRR_k are all irreducible asc-sets. Then we have the following.

MTP-PRINCIPLE Ⅱ (Strong Form). For a theorem T with hyp-set HYP and conc-pol $CONC$ to be true on the irreducible

component Var(IRR_k) of the variety Zero(HYP), it is necessary and sufficient that the remainder of $CONC$ w.r.t. IRR_k is 0:

$$\text{Remdr}(CONC/IRR_k) = 0.$$

The last principle MTP-Ⅱ is a complete one in that it can be completely decided whether a theorem is true or not true on this or that If the theorem T is generically true under the nondegeneracy conditions $I_i \neq 0$ so that in particular Z is a zero of $CONC$ by the formula above, then it follows from the theorem about generic zero in Sect 3 that we have necessarily $R = 0$ as a pol. Hence we have:

Principle of Mechanical Theorem Proving or MTP-Principle Ⅰ (Weak Form). If the remainder R of $CONC$ w.r.t. a char-set CS of HYP is 0, then the theorem T with hyp-set HYP and conc-pol $CONC$ is generically true under the non-degeneracy conditions $I_i \neq 0$ for which I_i are the initials of pols in CS. Moreover, if the char-set CS is irreducible, then the above condition $R = 0$ is also a necessary one for the theorem T to be generically true under the above non-degeneracy conditions.

Let us now apply the zero-decomposition theorem (weak form) (Ⅱ) to the hyp-set HYP so that

$$\text{Zero}(HYP) = \sum_k \text{Zero}(CS_k/J_k),$$

in which each CS_k is an asc-set and J_k is the product of initials of pols in CS_k. Let R_k be the remainder of $CONC$ w.r.t. CS_k. Then we have the following.

MTP-PRINCIPLE Ⅰ'(General Weak Form). if $R_k = 0$, then the theorem T is true on the part Zero(CS_k/J_k) of the whole variety Zero(HYP) of geometrical configurations verifying the hyp-set HYP. The condition $R_k = 0$ is also necessary if CS_k is irre-

ducible.

More generally, let us apply the variety-decomposition theorem to HYP so that we have

$$\mathrm{Zero}(HYP) = \sum_k \mathrm{V}(IRR_k),$$

in which IRR_k are all irreducible asc-sets. Then we have the following

MTP-PRINCIPLE Ⅱ (Strong Form). For a theorem T with hyp-set HYP and conc-pol $CONC$ to be true on the irreducible component $\mathrm{Var}(IRR_k)$ of the variety $\mathrm{Zero}(HYP)$, it is necessary and sufficient that the remainder of $CONC$ w.r.t. IRR_k is **0**:

$$\mathrm{Remdr}(CONC/IRR_k) = 0.$$

The last principle MTP-Ⅱ is a complete one in that it can be completely decided whether a theorem is true or not true on this or that irreducible component of the whole variety of Zero (HYP). No more degeneracy conditions will be intervened. As an example we may cite the Feuerbach Theorem stating that the 9-point circle of a triangle is tangent to the 4 inscribed and escribed circles of that triangle. With lines and circles oriented we can establish the hyp-set HYP with Zero(HYP) decomposed into 8 irreducible components. The conclusion is seen to be true just on 4 of such components which reflects the geometrical fact that each of the two oriented 9-point circles is tangent in the same sense just to two of the four oriented in-or es-cribed circles. For this or other examples we refer to the paper [Wu15]. we should remark that, in practice the less general MTP-Principle I seems to be preferred which is already so efficient that a lot of non-trivial and difficult theorems have already been proved and even discovered.

The mechanization methods as given above for elementary

geometries have also been extended to the differential case so that differentialgeometrial theorems can be proved also in a mechanical way as in the case of elementary geometries. We refer again to the relevant papers of Wu, particularly [Wu20]

5 Further Developments and Applications

In the Institute of Systems Science, Academia Sinica, there was organized a small MM-Group (Mathematics-Mechanization Group) devoting mainly to the developments and applications of the mechanization method of mathematics as explained in previous sections and other intimately related subjects. The group edited also an irregular publication "MM – Res. Preprints" for which 2 issues have since already appeared and the third one is also in the press. We list below some of such studies in these years.

A1. Mechanical Theorem Proving and Discovering of Elementary Geometries

This was the beginning and also the most well-developed direction of the whole MM-method (mathematics mechanization method). It was proved by Wu in starting from axiomatic systems that theorems of various kinds of elementary geometries, whether ordinary, affine, projective, non-euclidean, line or circle ones, etc., can all be proved in a mechanical manner by the general method, so far no order relations are involved, cf. [Wu3]. Such results are formulated in the form of Mechanization theorem of this or that geometry, just like the Hilbert Mechanical Theorem in Sect 2. Gao has also made a detailed comparative study of various geometries

connected with the parallelism axiom, cf [Gao3] For particular theorems we may mention the list of theorems proved or discovered by Wang and Gao as outlined in their joint paper [W - G]. The most extensive study has been undertaken by Chou and a book devoting to it has already been published, cf [Ch1, 2]. Furthermore, Wu has introduced the notion of oriented lines and circles to avoid or alleviate the difficulties of factorization in the case of ordinary geometry which has further increased the efficiency of the method and promise to discover much more difficult new theorems, cf. [Wu15]. Besides, some theorems in algebraic geometry and planar kinematics, have also been proved or discovered by Wu, of. [Wu19]. In the case of ordinary geometry for which the hypothesis set is already in the well-ordered form, Hong and Tau-Zhang-Yang have proposed very interesting methods of mechanical theorem proving by means of a single example or a set of examples. cf. e. g. [H1, 2]and papers of Tau et al to be published.

A2. Mechanical Theorem Proving and Discovering of Differential Geometries

In principle for differential geometries the same can be done as in the case of elementary geometries but much less have been actually carried out. For concrete results we may cite a detailed study of theorems about pairs of curves of Bertrand type in either ordinary or affine space, cf [Wu14]. The same has also been partially carried out including a theorem generalizing the classical Baecklund theorem to the affine space which has been considered in recent years by S. Chern et al. (to be published).

A3. Automatic Determination of Unknown Relations

Wu showed how by well-arranging the variables involved our method will give rise automatically to the desired unknown relations. One of the simplest of such examples is furnished e. g. by the automatic discovery of Chin-Heron formula of the area of a triangle in terms of its 3 sides, cf. [Wu12, 13]. It seems that the same method can furnish us with a formula for the volume of a hyperbolic tetrahedron in terms of its 6 edges, a problem which has greatly attracted the attention of mathematicians in recent years. Note that the occurrence of highly transcendental functions is not an insurmountable barrier for applying out method, see A6 below.

A4. Automatic Loci Determination

Again by well-arranging the variables involved our method will permit to give automatically the equations of loci to be studied. A concrete example is furnished by the determination of coupler-curve equation of a 4-bar-linkage by Wu (unpublished).

A5. Problems about Polynomials

Hu and Wang have discovered a new method of factoring multivariate polynomials with coefficients in an arbitrary finite extension field of rational number field which is based on our general MM-method, cf. [H-W]. Gao has introduced the notion of general discriminant system of a unary polynomial and gave also an algorithm of its determination which has been improved by W. D. Wu, cf. [Gaol] and [WDWu]. Liu has studied the com-

plex root isolation and degree-depressing of polynomials by decomposition instead of factorization, cf. [Liu1, 2]. Li has discovered an interesting general phenomenon of occurrence of unavoidable factors which generalizes the known theorem of Collins about successive pseudo-divisions of two polynomials, cf. [LiI-3]. In connection with such phenomenon Wu has proposed some alternative procedures in elimination for the formation of char-sets which would avoid the occurrence of such un-desirable factors, cf. [Wu10]. Concrete examples have shown that efficiency of computations will be raised to a much higher level and our new implementation of programming has taken into account of this advantage.

A6. Problems involving Transcendental Functions

At first sight it seems that the occurrence of transcendental functions would be an insurmountable barrier to the application of our general MM-method which is essentially algebraic in character. However, as already pointed out by Wu since his earliest writings, quite often it is rather the algebraic relations among the transcendental functions, but not the functions themselves, which will play an actual role in the problems to be solved. The transcendence may then be rendered disappear if one replaces the transcendental functions by new variables connected by algebraic relations or by applying rationalization process. It is for this reason that a lot of theorems involving trigonometric or hyperbolic functions, including trigonometric identities, have been, proved or discovered by our MM-method. Recently Wu et al have shown how transcendental equations like the Kepler equation in astrono-

my can be turned into an algebraic (more exactly an algebrico-differential) ones so that one can solve it in an algebraic manner, in contrast to the generally used graphical or numerical methods of solution, cf. [W-W]. The method consists in replacing the transcendental function in question by a new function and a new differential equation determining that function. The method is again a general one and will be useful in many occassions.

A7. Studies on Limit Cycles

Some papers of Wang, e. g. [Wang1] have shown how our MM-method can be applied to study the Lyapounov functions at a singular point of a polynomial differential equation system in the plane. The Soviet and Chinese mathematicians have shown how such knowledge could be applied to the construction of limit cycles, whose study is seemingly the only one of the 23 problems of Hilbert which has little advance up to the present time.

A8. Automatic Derivation of Newton's Gravitational Laws from Kepler's Laws

In [Wu17] Wu has shown how Newton's gravitational laws can be automatically derived from the Kepler's Laws. This may serve as a simplest example of automatically formulating underlying mathematical principles from experimental facts in physical sciences.

A9. Inequalities Proving

A lot of inequalities occurring in elementary mathematics can be put into the form of equations-solving by means of method of Lagrange's multipliers. The latter one can then be dealt with by our

general mechanization method of equations-solving. In [Wu18] Wu has shown how this can actually be carried out for various kinds of inequalities. There is also a paper of Wang [Wang2] dealing with the same subject.

A10. Non-linear Programming

The linear programming deals with the optimization of a linear function involving variables under linear constraints for which the variables and constraints are both large in number. On the contrary, the non-linear programming deals with optimization of a non-linear function involving variables under non-linear constraints for which the variables and constraints are both relatively few in numbers. The fewness of the number of constraints may be taken into great advantage by our method of inequalities-proving in A9 while the non-linearity of the function to be optimized and the constraints may be treated in the usual manner by our method. Wu has thus been able to solve in a new way certain problems in chemical engineering, viz. chemical equilibrium problem and chemical reaction optimization problem (to be published). Note that transcendental functions occurring quite often in non-linear programming can be dealt with again by the tactics in A6.

A11. Robotics

The solving of inverse kinematical equations (I. K. E.) of robot motions is usually considered as one of the most difficult problems in robotics studies. However, Wu has applied his MM-method of equations-solving to give simple solutions of I. K. E.

for certain simple robots usually used in practice and is able to furnish the equations of continuous variation of control parameters meeting the continuous motion of robot hands (to be published). A complete solution in the general case is quite difficult but seems to be still possible and is yet in the course of study.

A12. Mechanical Design and Control Theory

In some early papers in Chinese Wu has indicated how our mechanization method of equations-solving can be used to some simple problems of linkage design, pole-assignment of control theory, etc. Now in the subsequent years the programming has been greatly improved and the MM-Group has been equipped with much more powerful computers than the ones available at that time. It seems that much harder problems in these realms will be successfully attacked in years to come.

References

[BL] Bledsoe, W. W. and Loveland, D. W. (Ed.), *Automated Theorem Proving, after 25 Years*, Amer. Math. Soc., (1984).

[CH1] Chou, S. S., Proving elementary geometry theorems using Wu's algorithm, in [BL], 243—286.

[CH2] —, *Mechanical geometry theorem-proving*, Reidel, (1988).

[Gao1] Gao Xiao-shan, The discriminant systems of unary equations and their computation, *MM-Res. Preprints*, No. 1 (1987) 13—32.

[Gao2] —, Transcendental functions and mechanical theorem proving in elementary geometries, *MM-Res. Preprints*, No. 2 (1987) 37—47.

[Gao3] —, Mechanical theorem proving in Cayley-Klein geometries, *MM-Res. Preprints*, No. 3 (1989).

[Gao4] —, The minimal characteristic set of polynomial ideal, *J. Sys. Sci. and Math. Scis.*, 2 (1989).

[H1] Hong Jiawei, Can we prove geometry theorem by computing an example? *Scientia Sinica*, A29 (1986) 824—834.

[H2] —, How fast the algebraic approximation can be? *ibid*, A29 (1986) 813—823.

[H-W] Hu Sen and Wang Dongming, Fast factorization of polynomials over rational number field of its extension fields, *Kuxue Tongbao*, 31 (1986) 150—156.

[J-W] Jin Xiaofan and Wang Dongming, On Kukles' condition for the existence of a center, *MM-Res. Preprints*, No. 3 (1989).

[Li1] Li Zi-ming, A new proof of Collins's theorem, *MM-Res. Preprints*, NO. 1 (1987) 33—37.

[Li2] —, On the trianguration for any finite polynomial set (I), *MM-Res. Preprints*, No. 2 (1987) 48—54.

[Li3] —, Determinant polynomial sequences, *MM-Res. Preprints*, No. 3, (1989).

[Liu1] Liu Zhuo-jun, A method of isolating complex roots of polynomials—An application of Wu-Ritt Principle, *MM-Res. Preprints*, No. 2 (1989) 55—61.

[Liu2] —, Algorithm of decomposing high degree polynomials, *MM-Res. Preprints*, NO. 2 (1989) 62—67. also *Chinese Q. J. Math.* 3 (1988) 101—108.

[Ritt1] Ritt, J. F., *Differential equations from the algebraic standpoint*, Amer. Math. Soc., (1932).

[Ritt2] —, *Differential algebra*, Amer. Math. Soc., (1950).

[Wang1] Wang Dong-ming, Mechanical manipulation of differential systems, *MM-Res. Preprints*, NO. 1 (1987) 38—52.

[Wang2] —, A decision method for definite polynomial, *MM-Res. Preprints*, NO. 2 (1987) 68—74.

[Wang3] —, On Wu's method for proving constructive geometric theorems,

MM-Res. Preprints, NO. 3 (1989).

[Wang4] —, Proving-by-examples method and inclusion of varieties, *Kexue Tongbao*, 33 (1988) 2015—2018.

[W-G] Wang Dong-ming and Gao Xiao-shan, Geometry theorems proved mechanically using Wu's method—Part on Euclidean geometry, in *MM-Res Preprints*, NO. 2, (1987) 75—106.

[WDWu] Wu Wen-da, A note to the discriminant systems for the unary equations, *MM-Res. Preprints*, No. 3 (1989).

[Wu1] Wu Wen-tsun, On the decision problem and the mechanization of theorem-proving in elementary geometry, *Scientia Sinica* 21 (1978) 159—172, re-published in [BL], 213—234.

[Wu2] —, Some recent advance in mechanical theorem-proving of geometries, in [BL], 235—242.

[Wu3] —, *Basic Principles of Mechanical Theorem Proving in Geometries (Part on Elementary Geometries)*, (in Chinese), Science Press, (1984).

[Wu4] —, Basic principles of mechanical theorem-proving in elementary geometries, *J. Sys. Sci. and Math. Scis.*, 4 (1984) 207—235. Republished in *J. of Automated Reasoning*, 2 (1986) 221—252.

[Wu5] —, Toward mechanization of geometry—Some comments on Hilbert's "Grundlagen der Geometrie", *Acta Math. Scientia*, 2 (1982) 125—138.

[Wu6] —, Some remarks on mechanical theorem-proving in elementary geometry, *Acta Math. Scientia* 3 (1983) 357—360.

[Wu7] —, On the mechanization of theorem-proving in elementary differential geometry, *Scientia Sinica*, *Math. Supplement* (*I*), 94—102 (1979). (in Chinese).

[Wu8] —, Mechanical theorem proving in elementary geometry and differential geometry, in *Proc. 1980 Beijing DD-Symposium*, Beijing, v. 2, 1073—1092 (1982).

[Wu9] —, A constructive theory of differential algebraic geometry, in Dif-

ferential Geometry and Differential Equations, *Lect. Notes in Math*. No. 1255, *Springer* (1984) 173—189.

[Wu10] —, On zeros of algebraic equations —an application of Ritt principle, *Kewue Tongbao* 31 (1986) 1—5.

[Wu11] —, A mechanization method of geometry I. Elementary geometry, *Chinese Q. J. Math.*, 1 (1986) 1—14, Errata and addenda, *ibid*, 1 (1987) 20.

[Wu12] —, A mechanization method of geometry and its applications, I. Distances, areas, and volumes, *J. Sys. Sci. and Math. Scis.*, 6 (1986) 204—216.

[Wu13] —, A mechanization method of geometry and its applications, I. Distances, areas, and volumes in euclidean and non-euclidean geometries, *Kuxue Tongbao* 32 (1986) 436—440.

[Wu14] —, A mechanization method of geometry and its applications, 2. Curve pairs of Bertrand type, *Kuxue Tongbao* 32 (1987) 585—588.

[Wu15] —, On reducibility problem in mechanical theorem proving of elementary geometries, *Chinese Quarterly J. of Math.*, 2 (1987) also in *MM-Res. Preprints*, NO. 2, (1987) 18—36.

[Wu16] —, A zero structure theorem for polynomial-equations-solving and its applications, *MM-Res. Preprints*, No. 1 (1987) 2—12.

[Wu17] —, Mechanical derivation of Newton's Gravitational Laws from Kepler's Laws, *MM-Res. Preprints*, No. 1, (1987) 53—61.

[Wu18] —, A mechanization method of geometry and its applications, 3. Mechanical proving of polynomial inequalities and equations-solving, *MM-Res. Preprints*, No. 2, (1987) 1—17, also in *J. Sys. Sci. and Math. Scis.*, Inst. of Systems Science, 2 (1989).

[Wu19] —, A mechanization method of geometry and its applications, 4. Some theorems in planar kinematics, *J. Sys. Sci. and Math. Scis.*, Inst. of SYstems Science, 2 (1989) 97—109.

[Wu20] —, On the foundation of algebraic differential geometry, *MM-*

Res. Preprints, No. 3 (1989), also *J. Sys. Sci. and Math. Scis.*

[Wu21] —, Some remarks on characteristic-set formation, *MM-res. Preprints*, No. 3 (1989).

[W-W] Wu Wen-tsun and Wu Tianjiao, A mechanization method of geometry and its applications, 5. Solving transcendental equations by algebraic methods, *MM-Res. Preprints*, No. 3 (1989).

On the development of polynomial equations solving in China*

1 Introduction.

In Chap. 2 of the classic 《Mathematical Discovery》 Polya has called our attention about a posthumous work of Descartes, 《Rules for the Direction of the Mind》. Polya described a rough outline of the scheme that Descartes expected to be applicable to all types of problems which we reproduce below:

First, reduce any kind of preblem to a mathematical problem.

Second, reduce any kind of a mathematical problem to a problem of algebra.

Third, reduce any problem of algebra to the solution of a single equation.

The third step may be further clarifed as shown in the paragraph below which we again reproduce from the above classic of Polya:

Descartes advises us to set up as many equations as there are

* Proc. 1991 Nankai Seminar on Mathematics Mechanization, or Computer Mathematics (Eds. W. T. Wu & G. D. Hu), Singatore, (1993) 1—7

unknowns. Let n stand for the number of unknowns, and $x_1, x_2, \cdots, x_n$ for the unknowns themselves; then we can write the desired system of equations in the form

$$r_1(x_1, x_2, \cdots, x_n) = 0,$$
$$r_2(x_1, x_2, \cdots, x_n) = 0,$$
$$\cdots\cdots$$
$$r_n(x_1, x_2, \cdots, x_n) = 0,$$

where $r_1\ (x_1, x_2, \cdots, x_n)$ indicates a polynomial in $x_1, x_2, \cdots, x_n$, and the left-hand sides of the following equations must be similarly interpreted. Descartes advises us further to reduce this system of equations to one final equation.

It seems that Descartes was absolutely absurd in reducing any kind of problem to one of polynomial equations solving. Even the last step of reducing such a system of polynomial equations to one single final equation is not at all evident how to do and is actually entirely non-trivial. However, let us reproduce again some paragraph from the same classic:

> Descartes' project failed, but it was a great project and even in its failure it influenced science much more than a thousand and one little projects which happened to succeed. Although Descartes' scheme does not work in all cases, it does work in an inexhaustible variety of cases, among which there is an inexhaustible variety of *important* cases.

In fact, in now-a-days we see that there are really inexhaustive problems from mathematics, from sciences, from technologies, etc. which will ultimately be reduced to polynomial equations solving. Among all kinds of systems of equations to be

solved, those of polynomial forms are clearly the simplest, the most fundamental, and at the same time entirely non-trivial one. Among all kinds of problems in mathematics, the problem of polynomial equations solving should therefore be considered as one of primary importance to be settled with highest priority.

2 Polynomial Equations Solving in Ancient China.

In contrast to the Euclidean system of ancient Greek for which theorem-proving was the main concern, equations-solving occupied the central position in Chinese ancient mathematics. Below we shall give a brief outline of the main achievements of our ancestors on this subject.

The earliest mathematical classic of ancient China which fortunately had been preserved up to present day was 《*Nine Chapters in Arithmetic*》 (abbr. *NC*), completed about 1c B. C. In that classic there had already appeared the method of solving a system of linear equations together with the introduction of negative numbers which is actually the same as the present-day Gaussian elimination method. The coefficients of linear equations were represented by well-arranged counting rods put in a matrix form on a counting board. By manipulating the counting rods in some mechanical way the matrix of coefficients will be reduced to a form which in modern notations is in general equivalent to a set of equations as shown below:

$$c_{11} * x_1 + d_1 = 0,$$

$$c_{22} * x_2 + c_{21} * x_1 + d_2 = 0,$$

$$\cdots \qquad (2.1)$$

$$c_{nn} * x_n + \cdots + c_{n1} * x_1 + d_n = 0,$$

In the above equations (eventually some may be lacking as shown by one example in *NC*) all the leading coefficients c_{ii} are non-zero and the solutions are then readily found (eventually in terms of some parameters corresponding to the leading variables x_{ii} of the missing equations).

In another classic 《*Annotations to NC*》 (abbr. *AN*) appeared in 263A. D. by Liu Hui there were shown some variants of the above elimination method, comparison of efficiency of these variants, and detailed computations of some concrete equation-system in 5 unknowns.

There was already some trace of quadratic equations in *NC*. In fact, the extraction of square root corresponding to the solving of simplest quadratic equation $x^2 = a$ was solved by means of some geometrical principle, viz. the so-called *Out-In Complementary Principle*. This geometrical method of square-root extraction led to the solving of more general quadratic equations. This may be seen from most of our ancient classics that instead of saying as in our modern language *solving quadratic equation of the form* $x^2 + a * x = b$, *one said extracting square root of b with* cong a. In 《*NC*》 there was even a geometrical problem which was reduced to the square-root extraction with *cong*. In 《*AN*》 Liu Hui had given proofs of the validity of the above methods. In a classic of Zhao Shuang in about the same period as Liu Hui it was implicitely indicated how to express the solutions of a quadratic equation in terms of its coefficients. However, we have to remark that our ancestors seems to have little interest in the general expression of

roots of equations in terms of coefficients of the equations . They were rather interested in the determination of their real numerical solutions. This attitude of meeting practical applications permeated the whole period of later development of equations-solving in our ancient time.

The geometrical method of square-root extraction worked well also for cubic-root extraction corresponding to the solving of cubic equations of particular form $x^3 = a$, as already shown in 《*NC*》 as well as 《*AN*》. It was natural to extend the method to the solving of general cubic equations, as in the case of quadratic equations. Probably the celebrated mathematician Zu Congzhi of 5c had tried to solve such general cubic equations , though it was not quite certain since no written records were preserved. However, in the beginning of Tang Dynasty in 7c there appeared the classic of Wang Xiaotong in which were exhibited complicate problems arising from geometry and prac-tice. After the description of how the final cubic equations should be formed from such concrete problems, Wang gave directly the numerical solutions in saying simply *by root-extraction* .This shows that before the time of Wang the intellectuals should already know how to solve in numerical values a given cubic equation in numerical coefficients.

The way of further development was quite clear. Thus, according to the late renown historian in Chinese ancient mathematics Professor Yan Dongji, in calendars of Sung Dynasty (10c) there were biquadratic equations to be solved numerically. Unfortunately Professor Yan died prematurely in 1988 and no written manuscripts were left to us. In any way , throughout the whole period of Sung Dynasty (10-13c, very likely beginning from

much earlier time) efforts were exerted for the solving of polynomial equations of arbitrary degree. The goal was ultimately reached at latest in 13c. In fact, in 1249 there appeared the classic *Mathematics in Nine Chapters* of Qin Jiushao of Sung Dynasty. In that classic algorithm for the determination of numerical solutions of a polynomial equation of arbitrary degree was given which is actually the same as the much later re-discovered Horner 's method.

During the period of Sung and Yuan Dynasties there was a discovery of utmost importance in the introduction of the precise notion of an *unknown* under the name of *Heaven's Element*. A so-called *Heaven ' s Element Method* was developed which permitted to turn easily geometrical entities into the form of polynomials and geometrical relations into polynomial equations. It was known that there were texts about treat-ment of pairs of polynomial equations in two unknowns *Heaven's* and *Earth's Elements* but unfortunately such texts were lost. Fortunately there remained a classic *Four Ele-ments of Jade Mirrors* of Zhu Szejie in 1303 of Yuan Dynasty which treated polynomial equations up to even four unknowns. The method is one of elimination and may be seen clearly from one of the representative examples in Zhu's classic as shown below.

A geometrical problem was led to a set of three equations in three unknowns which in modern notations may be written in the following form (after removal of the factor x which will be supposed to be non-zero for simplicity):

$$x*y*z-x*y^2-z-x-y=0,$$
$$x*z-x^2-z-y+x=0,$$
$$z^2-x^2-y^2=0.$$

Zhu's method runs briefly as follows. Eliminate first y to get two equations in x and z alone. Eliminate then x to get a single equation in z alone. Choose from these equations three to be written in the modern form below:

$$z^4-6*z^3+4*z^2+6*z-5=0,$$
$$x^3-(z+2)*x^2+(z^2+4*z+2)*x-(2*z^2+2*z)=0. \tag{2.2}$$
$$y-(x*z-x^2-z+x)=0.$$

Solve the equations (2. 2) successively in single unknowns z, x, and y we get one of the solutions to be $(z, x, y) = (5, 3, 4)$. In general Qin Jiushao's method may be applied to get all possible real solutions if required.

3 Polynomial Equations-Solving in MMRC of Modern China.

Let K be a field of characteristic 0 and $x = (x_1, \cdots, x_n)$ be a set of indeterminates to be fixed in what follows. By a *pol* will be meant a polynomial in the ring $K[x_1, \cdots, x_n]$ and by a *polset* a finite set of such pols. For a polset PS consisting of non-zero pols $P_1, \cdots, P_n$ we shall write simply $PS=0$ for the set of equations $P_1=0, \cdots, P_n=0$. The totality of all possible solutions of $PS=0$ or zeros of PS in all conceivable extension fields of K will be denoted by *Zero* (PS). For any other pol G we shall write *Zero* (PS/G) for the set *Zero* (PS) -*Zero* (G), i. e.

the totality of zeros of PS which are not zeros of G. The problem of polynomial equations-solving amounts then to the determination of the set *Zero* (PS) for any given polset PS.

The equations (2.1) and (2.2) serve now as a clue for the solving of polynomial equations which had been carried out in recent years by the Mathematics-Mechanization Research Center (MMRC) of Institute of Systems Science in our Academy, cf. e. g. [WU1,2] .

To begin with, let us first introduce the following all the more important concept of an *ascending set* (abbr. *asc-set*) which we have borrowed from the two classics of J. F. Ritt, cf. [R1, 2].

For a non-constant pol P let c be the greatest subscript for which x_c appears actually in P. Let d be then the degree of P considered as a pol in x_c. Such a pol can then be written in the *canonical form*

$$P = I * x_c^d + \textit{lower degree terms in } x_c$$

with coefficients themselves pols in the ring $K[x_1, \cdots, x_{c-1}]$. We call c the *class* and d the *class-degree* or simply *degree* of P, to be denoted by *cls* (P) and *cdeg* (P) or simply *deg* (P) resp. The leading coefficient I *of* P w. r. t. the leading variable x_c is then called the *initial* of the pol P. For a non-zero constant pol the class will be defined to be 0 while the degree may be either defined as -1 or may be left undefined.

Consider now a polset AS of the following particular form. Suppose that the inde-terminates $x_1, \cdots, x_n$ can be rearranged as $u_1, \cdots, u_d, y_1, \cdots, y_r$ $(r + d = n)$ such that the pols in AS can be arranged as $A_1, \cdots, A_r$ in canonical forms below:

$$
\begin{aligned}
A_1 &= I_1 * y_1^{d_1} + \textit{lower degree terms in } y_1, \\
A_2 &= I_2 * y_2^{d_2} + \textit{lower degree terms in } y_2, \\
&\cdots\cdots \qquad (3.1) \\
A_r &= I_r * y_r^{d_r} + \textit{lower degree terms in } y_r .
\end{aligned}
$$

In (3.1) for each i the coefficients of y_i-powers in A_i are pols in $K[u_1, \cdots, u_d, y_1, \cdots, y_{i-1}]$. We suppose also each pol I_i has a degree in y_j less than $deg(A_j) = d_j$ for any pair (i, j) with $j < i$. In that case we say that the polset AS is an *ascending set* and the coefficient pols I_i are *initials* of the corresponding pols A_i. Note that in returning back to original notations, y_i are just x_{c_i} with $c_i = cls(A_i)$ where

$$0 < c_1 < \cdots < c_r,$$

while $u = (u_1, \cdots, u_d)$ are the remaining $x's$ arranged say in the same order as before.

mcomparing with the equations (2. 1) and (2. 2) in Sect 1, we see clearly that an asc-set defined by (3. 1) is just a generalization of the polsets occuring in the left-side of (2. 1) and (2. 2).

A polset consisting of a single non-zero constant will be called a *trivial* asc-set. An asc-set of form (3. 1) will then be said to be *a non-trivial* one. It is clear that the zero-set of a trivial asc-set is empty, and for a non-trivial asc-set AS the zero-set Zero (AS) may also be considered as completely determined.

Our method of solving polynomial equations consists of now in reducing the problem of solving arbitrary polynomial equations to those of solving such equations correspond-ing to asc-sets. One of the variants of such method is expressed in the theorem below.

Main Theorem. There is an algorithm which permits to determine for any polset PS a finite set of asc-sets AS_i in a finite number of steps such that

$$\text{Zero}(PS) = \text{Sum}_i \text{Zero}(AS_i / IP_i), \tag{3.2}$$

in which each IP_i is the product of all initials of pols in the corresponding AS_i.

The formulas (3.2) as well as its variants will be called *decomposition formulas* for *Zero* (PS). The corresponding Main Theorems and the decomposition formulas form the corner-stones on which are based our method of polynomial equations-solving as well as all the applications. We leave these to the relavant papers published mostly in the irregular publication Mathematics Mechanization Research Preprints (abbr. MM-Preprints) .

REFERENCES

[R1] J. F. Ritt, Differential equations from the algebraic standpoint, Amer. Math. Soc., Providence (1932).

[R2] J. F. Ritt , Differential algebra, Amer. Math. Soc. Providence (1950).

[WU1] Wu Wen-tsun, On Zeros of algebraic equations—an application of Ritt principle, Kexue Tongbao 31 (1986) 1-5.

[WU2] Wu Wen-tsun, A zero structrue theorem for polynomial-equations-solving and its applications, MM-Res. Preprints, No. 1 (1987) 2-12.

Polynomial equations – solving and its applications*

Abstract. We give first a rapid review of the development of polynomial equationssolving in ancient China. We show then how the achievements and lines of thought of our ancestors have been elaborated into a general method of solving arbitrary systems of polynomial equations with the aid of concepts and tools from modern mathematics. As illustrations of our general method, we give finally diverse applications to problems arising either from mathematics proper, or from other domains beyond mathematics, science and technology.

Let K be a field of characteristic 0 and $X = (x_1, \cdots, x_n)$ba a set of indeterminates to be fixed in what follows. By a *pol* will be meant a polynomial in the ring $K|X|$ and by a *polset* a finite set of such pols. For a polset PS consisting of non – zero pols $P_1, \cdots,$ p_m we shall write simply $PS = 0$ for the set of equations $P_1 = 0,$ $\cdots, P_m = 0$. The totality of all possible solutions of $PS = 0$ or zeros of PS in all conceivable extension fields of K will be denoted by Zero(PS). For any other pol G we shall write Zero(PS/G) for the set Zero(PS) \ Zero(G), i. e. the totality of zeros of PS

* Proc. ISSAC 94, or Algorithms and Computations, (Eds. D. Z. Du & X. S. Zhang), Springer(1994) 1—9.

which are not zeros of G.

By *polynomial equations – solving* is meant the determination of Zero(PS) itself or at least its structure for any polset PS. As variety of problems, either from mathematics proper or from other domains, will lead naturally to such polynomial equations $PS=0$, an effective method of such a polynomial equations – solving will have naturally an immense field of possible applications to problems from mathematics, science, and technology. Even the mechanical proving of geometry theorems, which is seemingly of a character entirely different from equations – solving, relies heavily on the study of the structure of Zero(HYP) with HYP the hypothesis polset of the geometry theorem in question. Besides, we may recall that the algebraic geometry, one of the most active field of reseach in modern mathematics, is in fact nothing else but the study of the structure and the properties of Zero(PS) for an arbitrary polset PS.

In view of the significance of polynomial equations – solving in both theory and practice as described above we shall exhibit in what follows the past achievements and present day developments of this subject restricted however mainly to our own country, China.

1 Polynomial Equations – Solving in Ancient China.

While Euclid's《*Elements*》in 3c B. C. laid down the foundation of a deductive system of geometry which dominates up to present day modern mathematics in the western world, the Chinese classic《*Nine Chapters in Arithmetic*》or《*Nine Chapters*》for

short in 1c B.C. had laid down the foundation of an entirely different mechanical system of mathematics which governed the development of Chinese mathematics up to Ming Dynasty in 14c A.D. While in Euclid's system theorem-proving is the main concern of mathematics, the mechanical system of ancient China emphasized equations – solving as the central theme of mathematics. Below we shall cite some of the main achievements of our ancestors on polynomial equations – solving.

In 《*Nine Chapters*》there had already appeared the method of solving a system of linear equations(together with the introduction of negative numbers)which is actually the same as the present-day Gaussian elimination method. The coefficients of linear equations were represented by well-arranged counting rods put in a matrix form on a counting board. By manipulating the counting rods in some mechanical way the matrix of coefficients will be reduced to a form which in modern notations is in general equivalent to a set of equations as shown below($c_{ii} \neq 0$):

$$c_{ii} * x_i + c_{i,i-1} * x_{i-1} + \cdots + c_{i1} * x_1 + d_i = 0, \ i = 1, \cdots, n. \tag{1.1}$$

In the last equations eventually some may be lacking as shown by one example in 《*Nine Chapters*》. In any way the solutions are readily found(eventually in terms of some parameters corresponding to the leading variables x_{ii} of the missing equations).

Numerical Solutions of Algebraic Equations in a Single Unknown.

Based on geometrical consideration the ancient Chinese knew in quite remote times extraction of square and cubic roots. As extensions of the root-extraction method it was also known how to

solve numerically a quadratic equation, as already shown in《*Nine Chapters*》. After hundreds of years of efforts there appeared ultimately in 1247 a classic《*Mathematics in Nine Chapters*》of Qin Jiushao in Song Dynasty. In that classic algorithm for the determination of numerical solutions of an algebraic equation of arbitrary degree was given which is actually the same as the later re-discovered Horner's method.

Method of Heaven's Element and Algebrization of Geometry.

During the period of Song and Yuan Dynasties there was a discovery of utmost importance in the introduction of the precise notion of an *unknown* under the name of *Heaven's Element*. A so-called *Heaven's Element Method* was developed which permitted to turn easily geometrical entities into the form of polynomials and geometrical relations into polynomial equations. This led to a more or less systematic reduction of geometrical problems to polynomial equations-solving and some kind of algebrization of geometry.

The Classic of Zhu Shijie about Method of Four Elements.

It was known that there were texts about treatment of pairs of polynomial equations in two unknowns *Heaven's* and *Earth's Elements* but unfortunately such texts were lost. Fortunately there remained a classic《*Jade Mirrors of Four Elements*》of Zhu Szejie in 1303 of Yuan Dynasty which treated polynomial equations up to even four unknowns. The method is one of elimination and may be seen clearly from one of the representative examples in Zhu's classic as shown below.

A Representative Example in Zhu's Classic.

A geometrical problem was led to a set of three equations in

three unknowns which in modern notations may be written in the following form:

$$x * y * z - x * y^2 - z - x - y = 0,$$
$$x * z - x^2 - z - y + x = 0,$$
$$z^2 - x^2 - y^2 = 0.$$

From geometrical conditions x cannot be zero and will thus be divided out freely during the procedure. Zhu's method runs now briefly as follows. Eliminate first y to get two equations in x and z alone. Eliminate then x to get a single equation in z alone. Choose from these equations three to be written in the modern form below:

$$z^4 - 6 * z^3 + 4 * z^2 + 6 * z - 5 = 0,$$
$$x^3 - (z + 2) * x^2 + (z^2 + 4 * z + 2) * x - (2 * z^2 + 2 * z) = 0,$$
$$y - (x * z - x^2 - z + x) = 0. \tag{1.2}$$

Solve the equations (1.2) successively in single unknowns z, x, and y we get one of the solutions to be $(z, x, y) = (5, 3, 4)$. In general Qin Jiushao's method may be applied to get all possible real solutions if required.

Conclusion.

For the actual computation one had to place the coefficients of various kinds of terms in counting rods at definite positions of the counting board. This limited the method to at most four equations in four unknowns and only quite simple ones can be so treated. However, the general character of the way to solve arbitrary set of polynomial equations is quite clear.

2 Guiding Principle and Methods of Polynomial Equations-Solving.

Systematization and Modernization of Zhu Shijie's Method.

In recent years the Mathematics-Mechanization Research Center(MMRC) of Institute of Systems Science in our Academy had developed a mechanization method of solving arbitrary set of polynomial equations in the form of determination of *Zero*(*PS*) for an arbitrary polset *PS*. The method, though the intervention of concepts and techniques in modern mathematics (mainly from Ritt, cf. [R1, 2]) is inevitable, is in the main along the lines of thought of our ancestors. It is in essence elaboration, modernization and systematization of the elimination methods of 《*Nine Chapters*》in the linear case and of Zhu Szejie in the general case. Our study is thus actually a continuation of our ancestor's works on equations-solving stopped since 14c. For literature we refer particularly to the relevant papers in the irregular publication Math. Mech. Res. Preprints (abbr. MM-Res. Preprints) of MMRC, cf. [MM].

Some General Concepts.

To begin with, let us first introduce the following all the more important concept of an *ascending set* (abbr. *asc-set*).

For a non-constant pol P let c be the greatest subscript for which x_c appears actually in P. Let d be then the degree of P considered as a pol in x_c. Such a pol can then be written in the *canonical form*

$$P = I * x_c^d + \textit{lower degree terms in } x_c.$$

With coefficients themselves pols in the ring $K[x_1, \cdots, x_{c-1}]$. We shall call c the *class* and d the *degree* of P, to be denoted by Cls(P)and Deg(P)resp. For a non-zero constant pol the class will be defined to be 0 while the degree will be left undefined.

Consider now a polset AS for which the pols in AS can be arranged as $A_1, \cdots A_r$ in canonical forms below:

$$A_i = I_i * x_{ci}^{di} + \textit{lower degree terms in } x_{c_i},\ i = 1, \cdots, r. \tag{2.1}$$

$$0 < c_1 < \cdots < c_r \leqslant n.$$

In(2.1)for each i the coefficients of x_{ci}-powers in A_i are pols in $K[x_1, \cdots, x_{ci-1}]$. We suppose also each pol I_i has a degree in xj less than $deg(A_j) = d_j$ for any pair$(i, j)$with $j < i$. In that case we say that the polset AS is an *ascending set* and the coefficient pols I_i are *initials* of the corresponding pols A_i.

In comparing with the equations(1.1)and (1.2)in Sect 1, we see clearly that an asc-set defined by(2.1)is just a generalization of the polsets occuring in the left-side of(1.1)and (1.2), with$(z, x, y) = (x_1, x_2, x_3)$in the latter case.

Given an asc-set defined by (2.1), a pol R for which the degree in x_{c_i} is less than d_i will be said to be ***reduced*** w.r.t. the pol A_i, and ***reduced*** w.r.t. the asc-set AS if this is so for each i.

Given an arbitrary pol G let us divide G successively by the pols in AS in the reverse order. This will give rise to an identity of the form

$$I_1^{s1} * \cdots I_1^{sr} * G = \sum i Q_i * A_i + R. \tag{2.3}$$

In the identity Q_i, R are pols with R reduced w.r.t. AS and s_i are non − negative integers which may be chosen to be as small as

possible. The pol R thus obtained from G will be called the *remainder* of G w.r.t. AS, to be denoted by

$$R = \text{Remdr}(G/AS).$$

The identity(2.3)is then called the *remainder formula* of G w.r.t. AS.

A polset consisting of a single non-zero constant will be called a *trivial asc-set*. An ascset of form(2.1)will then be said to be a *non − trivial* one. It is clear that $\text{Remdr}(G/AS) = 0$ for any pol G if the asc-set AS is a trivial one. For non-trivial asc-sets the notion of *irreducible* ones will be defined in a natural way.

Guiding Principle of Polynomial Equations-Solving.

For a non-trivial asc-set of form(2.1)we may solve $A_1 = 0$ for x_{c1} in numerical values when K is the rarional, the rael or the complex field and x_i, $i < c_1$, are given in numerical values. In general we may simply consider x_{c1} as a well-determined algebraic function of x_i, $i < c_1$, as defined by $A_1 = 0$. With x_{c1} thus determined we may then determine x_{c2} from $A_2 = 0$ either in numerical values or as a well-determined algebraic function in x_i, $i < c_2$, etc. In this sense Zero(AS) may thus be considered as well-determined. In case AS is a trivial asc-set, then it is clear that Zero(AS) is an empty set.

From the above we see that Zero(AS) and more generally Zero(AS/G)for any asc-set AS and any pol G may be considered as well determined. Our guiding principle for the solving of polynomial equations $PS = 0$ or equivalently the determination of Zero(PS) is to determine from the given polset PS a set of asc-sets so that Zero(PS)is closely connected or eventually completely determined from the zero-sets of these asc-sets.

General Methods of Polynomial Equations-Solving.

Following the guiding principle we get methods of polynomial equations-solving in the form of following theorems:

Main Theorem Ⅰ. There is an algorithm which permits to determine for any polset PS an asc-set CS in a finite number of steps such that

$$\text{Zero}(PS) \subset \text{Zero}(CS), \textit{ and} \tag{2.4}$$

$$\text{Zero}(PS) = \text{Zero}(CS/J) \cup \bigcup_i \text{Zero}(PS \cup \{I_i\}). \tag{2.5}$$

In(2.5) I_i are the initials of pols in CS and J is their product. The asc-set CS is to be called a *characteristic set* (abbr. char-set) of the polset PS.

Main Theorem Ⅱ. There is an algorithm which permits to determine for any polset PS a finite set of asc-sets AS_i in a finite number of steps such that

$$\text{Zero}(PS) = \bigcup_i \text{Zero}(AS_i/J_i), \tag{2.6}$$

in which each J_i is the product of all initials of pols in the corresponding AS_i.

Main Theorem Ⅲ. There is an algorithm which permits to determine for any polset PS a finite set of irreducible asc-sets IRR_k in a finite number of steps such that

$$\text{Zero}(PS) = \bigcup_i \text{Var}[IRR_k], \tag{2.7}$$

in which each $\text{Var}[IRR_k]$ is an irreducible algebraic variety well-determined from the irreducible asc-set IRR_k. Moreover, by mere computations one may remove away all redundant irreducible asc-sets so that in (2.7) none of the varieties in the union is contained in any one else.

Remark that (2.7) gives us a non-contractible decomposition

of the algebraic variety Zero(PS) in its irreducible components Var[IRR_k]which is unique up to an order.

The formulas(2.5), (2.6), (2.7) will be called *decomposition formulas* for Zero(PS). The Main Theorems and the decomposition formulas form the corner-stones on which are based our methods of polynomial equations-solving as well as all the applications. For general references cf. e. g. [Wu2]and some relevant papers of the author in[MM].

3 Applications of Polynomial Equations Solving.

3.1 Direct Applications to Polynomial Equations -Solving.

Many problems from mathematics or beyond mathematics are already in the form of polynomial equations solving. Let us cite below a few of them with references in [MM].

Problems of Enumerative Geometry.

Enumerative geometry asks for the number of certain kind of geometrical entities verifying some geometrical conditions which leads directly to the solving of some polynomial equations. It forms a beautiful application of modern algebraic geometry which deals with however only geometrical configurations in complex projective spaces and gives only the number of solutions in the *generic* case, being based on the so-called method of algebraic correspondence and the counting principle. Our method has no such restrictions. As an example, the Poncelet porism in case of triangles has been analyzed in some details by our method. Cf. e. g. [Wu5].

Problems from Linkages and Mechanisms.

The design problems of linkages and mechanisms lead often directly to the solving of polynomial equations. As examples we may cite the direct problem of Stewart platforms and the inverse kinematic equations of oobotic movements. For the former Wu Wenda et al have proved that the number of possible solutions (real as well as complex ones) of the corresponding equations is 40 in the generic case, while for the latter Liang et al have proved that the corresponding number of solutions in the generic case is 16. Some particular robots of Puma type, for which the neighboring joint axis are either parallel or orthogonal to each other have been studied some details by our methods.

Yang-Baxeter Equations from Statistical Physics.

The Yang-Baxeter equations in statistical physics in the case n=2 give rise to a system of 64 polynomial equations in 16 unknowns. Shi He has given a complete list of solutions of this system of equations by means of our methods. Among the solutions there are sets depending on several parameters which are unknown before. Besides, S. K. Wang, Institute of Applied Mathematics of our Academy, has treated Yang-Baxeter equations with spectrum parameters and discovered a number of entirely new solutions of physical interest by some elaboration of our methods.

3.2 Applications to Geometry Theorem Proving (MTP).

Apparently geometry theorem-proving (abbr. MTP) has little to do with polynomial equations-solving. That this is not so may be seen as follows.

For most of interesting theorems in geometries the hypothesis

and conclusion may be put in the form $HYP = 0$ and $CONC = 0$ respectively HYP being a polset and $CONC$ a pol in K $[X]$. We shall write such a theorem in the form $T = \{HYP, CONC\}$. Now a zero of Zero (HYP) is nothing else but a geometrical configuration verifying the hypothsis of the theorem which lies eventually in some extended space. Consequently if Zero (HYP) is known, then it will be relatively easy to ascertain on which part Z of Zero (HYP) the theorem will be true so that $Z \subset$ Zero ($CONC$). This is the general underlying principle of applying our methods of polynomial equations-solving to MTP. Remark in passing that for almost all known theorems of geometry it is *not ture* that Zero (HYP) $\subset$ Zero ($CONC$) so that some power of $CONC$ will be in the ideal of HYP, as a consequence of Hilbert's Nullstellen Satz. This shows that ideal-theoretical treatment seems to be inadequate to MTP of geometry. This is also the reason that we rely ourselves rather on the zeroset treatment for MTP of geometries as implied by the Main Theorems. Thus, for the decomposition formulas (2.5-7) let the polset PS be the hypothesis polset HYP of the theorem T in question. Then from the three Main Theorems we get respectively the following three MTP-Principles.

MTP-Principle I. For a given theorem $T = \{HYP, CONC\}$ let CS be a char-set of the hypothesis polset HYP. Let $R =$ Remdr ($CONC/CS$). Then $R = 0$ is a sufficient condition for theorem T to be *generically* true under the *non-degeneracy conditions* given by $J = 0$, or in other words Zero (HYP) $/J$) $\subset$Zero ($CONC$). Moreover, the condition $R = 0$ is also a necessary one for the above to be true in case CS is an irreducible asc

-set.

MTP-Principle Ⅱ. For a given theorem $T = \{HYP, CONC\}$ let AS_i be a set of asc-sets determined from the hypothesis polset HYP such that we have (2.6) with $PS = HYP$. Then the theorem T will be true on that part Zero (AS_i/J_i) of Zero (HYP) or Zero $(AS_i/J_i) \subset$ Zero $(CONC)$ if Remdr $(CONC/AS_i) = 0$.

MTP-Principle Ⅲ. For a given theorem $T = \{HYP, CONC\}$ let IRR_K be a set of irreducible asc-sets determined from the hypothesis polset HYP such that we have (2.7) with $PS = HYP$. Then the theorem T will be true on that irreducible component Var $[IRR_K]$ if and only if Remdr $(CONC/IRR_K) = 0$.

Remark that the last Principle Ⅲ furnishes us the most complete information about the truth of a theorem. However, in view of the complications for a complete decomposition of the form (2.7) with $PS = HYP$, it is often preferred to apply the much less complete MTP-Principle I. In fact, the mere application of this incomplete principle had already permitted to prove or even discover hundreds of non-trivial theorems in various kinds of geometries. Cf. for more informations e. g. the works [Wul, 4], [C], and papers of D. M. Wang, X. S. Gao, and the author in [MM].

3.3 Applications to Problems Reducible to Polynomial Equations-Solving.

There are lot of problems which are not in the form of polynomial equations-solving, can however by some devices be reduced to such ones. We shall state a few of them which we have dealt with.

D. M. Wang has applied our method of polynomial equations-solving to the qualitative study of polynomial planar system of ordinary differential equations. In particular, he has discovered some criteria for center and foci of quadratic or cubic systems and has thus pointed out some fallacies of well-known results of Kukles and others. It is also applied by Wang to fild out 6 limit cycles of some cubic system. Cf. e. g. [Wa] and also papers in [MM].

In the realm of partial differential equations Z. B. Li et al have studied for example the Belousov-Zhabotinskii (B-Z) reaction-diffusion equations. By setting the unknown functions as a definite type rational function of the exponential of a new unknown function, one gets a system of polynomial equations in the yet to be determined coefficients in the rational function. Applying our methods for the solving of this system of equations, it is found a complete set of travelling -wave solutions of the B-Z equations. In a similar way Li and others have found lot of travelling-wave solutions of various other well-known partial differential equations in mathematical physics which have been vividly studied in recent decades. Cf. e. g. [L-S].

Optimization and Related Problems.

Let us consider in the real n - space a domain D defined by

$$a_i \leq x_i \leq b_i i, \quad i = 1, \cdots, n,$$

in which a_i may be either finite or $-\infty$, and b_i either finite or $+\infty$. Let $f \in R[X]$, and $HS \subset R[X]$ be given.

Optimization Problem O. Determine the greatest and/or least values, if exist (e. g. when all a_i, b_i are finite), of f over D under the constraint conditions $HS = 0$.

The usual approach to such kind of problems by Lagrange multiplier method has long been forsaken owing to the fact that it will lead to the solving of equations difficult to deal with However, so far *HS* are resticated to polynomial ones, then our methods of polynomial equations-solving have revived this approach. In fact, a modified form of our Main Theorem II has led to the following

Finite Kernel Theorem. There is an algorithm to get a *finite* set K of real values such that the greatest and/or the least values of the Problem O, supposed exist, are given by the greatest and/or the least values of this finite set, to be called a ***kernel set*** of the Problem O.

Remark that the finiteness of the kernel set K comes just from the restriction that f and *HS* are all polynomial ones.

Many problems of non-linear programing may be solved by means of the above theorem. In particular, a collision problem of traffic has been studied in some details. Moreover, a lot of inequalities, algebraic, geometric, or trigonometric, have been proved in this way with relative ease. Cf. e. g. [Wu5] and the papers of T. J. Wu and the author in [MM].

For applications of our method to mechanical geometry theorem-proving we have deliberately avoided the order relations which algebraically will be turned into inequalities. It is an important problem to extend the mechanical proving to include theorems involving order relations. Our method as described above turns out to be quite efficient and a number of non-trivial theorems of such kind have already been proved in this way.

In CAGD we encounter the following

Problem GD. Determine in the real space an algebraic surface S of given degree to pass through a set of given algebraic curves and touch a set of given algebraic surfaces along given algebraic curves smoothly up to certain degree.

Based on algebraic geometry the above problem has been reduced to the solving of polynomial equations in the unknowns which are the coefficients of the algebraic surface S to be determined. As a simplest example, we may consider a pair of cylinders C_1, C_2 of radii r_1, $r_2$along X-and Y-axis and bounded by circles orthogonal to the X-and Y-axis and at distances d_1, d_2 from the origin respectively. Our problem is to determine cubic algebraic surfaces, if exist, to touch smoothly the cylinders along the bounding circles. Our methods of polynomial equations-solving show that for this to be possible it is necessary and sufficient that $r_1^2 + d_1^2 = r_2^2 + d_2^2$ and in that case there will be one and only one such cubic surface whose equation may be explicitely written down.

3. 4 Diverse Applications.

There are also problems which may be solved not by reducing to some problems of polynomial equations-solving, but by methods which occur in our methods of polynomial equations-solving.

Automated Discovering of Unknown Relations.

Problem AD. Suppose there are some algebraic and/or geometric entities for which we know that there should exist some relations between them. To discover the explicit form of such relations.

Our method of solving this kind of problems is briefly as fol-

lows. Let us designate the algebraic or geometric entities involved in the problem by $X = (x_1, \cdots, x_n)$ with the first few x's as those between which is to be determined the yet unknown relation. Let PS be the polset such that $PS = 0$ is the set of relations among the entities X. Form now a char-set CS of PS which the first polis C_1. By (2.4) we have Zero $(PS) \subset$ Zero $(CS) \subset$ Zero (C_1). Then $C_1 = 0$ will be the relation between the first few entities as required.

As a simplest example let us determine the area of a triangle in terms of its sides. In accordance to our general method let us denote the 3 sides of the triangle by x_1, x_2, x_3 and the area by x_4, while the 3 vertices by $(0, 0)$, $(x_5, 0)$, (x_6, x_7). Form the polset PS and a char-set CS of PS with first pol C_1. Then $C_1 = 0$ is nothing else but the wellknown Heron-Qin formula which has now been discovered in an automatic way. For more examples we may refer to [Wu3] and a paper of X. S. Gao and D. K. Wang in [MM].

Automatic Determination of Geometrical Loci.

The equations of geometrical loci may also be automatically determined by the above method. Consider e. g. a point moving in the plane under some geometrical constraints. Take coordinates with moving point as (x_1, x_2) and denote by x_i $(i > 2)$ the other geometrical entities intervening in the geometrical constraints. Write down the geometrical conditions in the form of $PS = 0$ and determine a char-set CS of PS. Then in general the first pol C_1 in CS will give the equation of the locus to be found as $C_1 = 0$. As a particular example, we may determine the coupler equation of a coupler point of a 4-bar linkage in this way. The re-

sult is automatically found, as is well known, to be a sextic equation in the coordinates $(x, y) = (x_1, x_2)$ of the coupler point.

Applications of our methods of polynomial equations-solving are steadily growing. Moreover it has been extended to differential case. The results are parallel to those of Sect 2 in replacing naturally pols and polsets by differential ones and the procedure and reasonings are naturally much more complicate. Applications to mechanical theorem proving of differential geometries and automatic discovering have also been done. In any way much has to be studied yet, in both ordinary and differential cases.

REFERENCES.

[C] S. C. Chou, Mechanical geometry theorem-proving, Reidel (1988).

[L-S] Z. B. Li & H. Shi, Exact solutions for Belousov-Zhabotinskii reaction diffusion system, J. Phys. A: Math. Gen, to appear.

[MM] MM-Res. Preprints, No. 1-11, (1985-1994).

[R1] J. F. Ritt, Differential equations from the algebraic standpoint, Amer. Math. Soc., Providence (1932).

[R2] J. F. Ritt, Differential algebra, Amer. Math. Soc., Providence (1950).

[Wa] D. M. Wang, Mechanical approach for polynomial set and its related fields, Doctoral Thesis, Institute of Systems Science, Acad. Sinica (1987).

[Wu1] Wu Wen-tsun, Basic Principles of Mechanical Theorem Proving in Geometries (Part on Elementary Geometries),

(in Chinese), Science Press, Beijing (1984), English translation by D. M. Wang et al, Springer (1994).

[Wu2] Wu Wen-tsun, On zeros of algebraic equations-an application of Ritt principle, Kexue Tongbao 31 (1986) 1-5.

[Wu3] Wu Wen-tsun, A mechanization method of geometry and its applications, I. Distances, areas, and volumes in euclidean and non-euclidean geometries, Kuxue Tongbao 32 (1986) 436-440.

[Wu4] Wu Wen-tsun, A report on mechanical geometry theorem proving, Progress in Natural Science, 2 (1992) 1-17.

[Wu5] Wu Wen-tsun, A mechanization method of equations-solving and theoremproving, Adv. in Comp. Res., 6 (1992) 103-138.